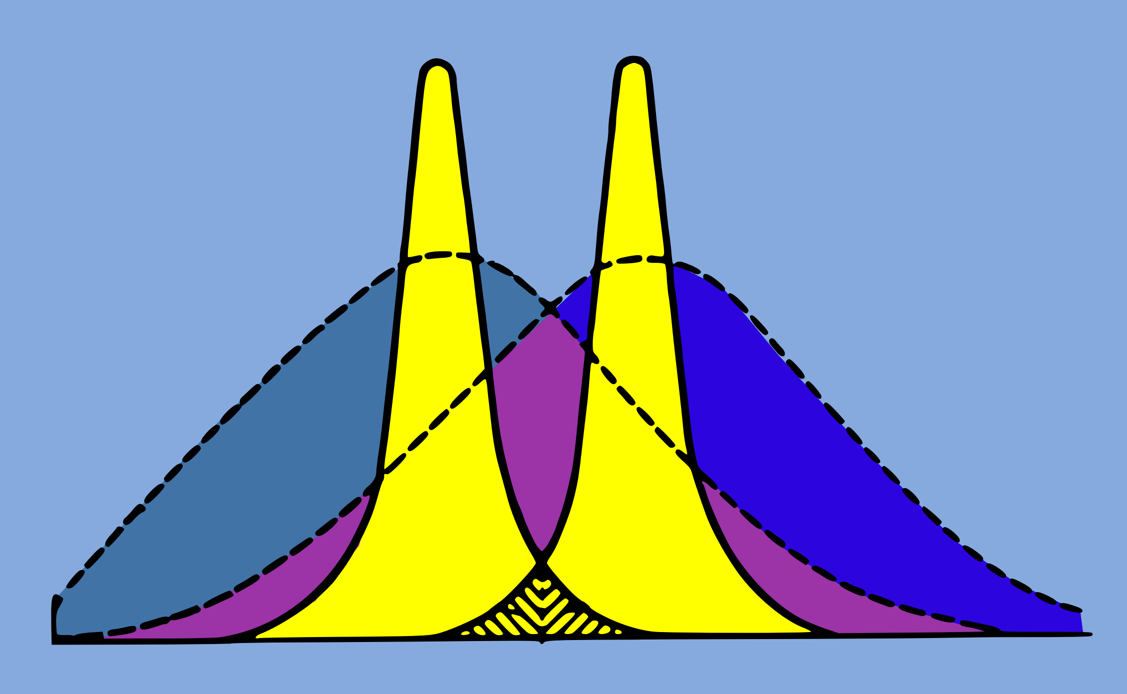
B. Lévine

Fondements Théoriques de la RADIOTECHNIQUE STATISTIQUE

Tome2



Éditions Mir Moscou

Б. ЛЕВИН

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СТАТИСТИЧЕСКОЙ РАДИОТЕХНИКИ

Том II

СОВЕТСКОЕ РАДИО • МОСКВА

B. LÉVINE

FONDEMENTS THÉORIQUES DE LA RADIOTECHNIQUE STATISTIQUE

Tome
II

ÉDITIONS MIR · MOSCOU

CDU = 621.39.1:519.27 = 40

Traduit du russe
par A. SOKOVA

на французском языке

Copyright by les Editions Mir
U.R.S.S. 1973

Л $\frac{0341\text{-}261}{041(01)\text{-}73}$

PRÉFACE DE L'AUTEUR À L'ÉDITION FRANÇAISE

Ce second volume paraît en même temps que le premier car, réunis sous un même titre, ces deux volumes constituent un tout. Les principes de base qui nous ont guidés dans la présentation du matériel ont déjà été exposés dans le chapitre d'introduction du premier volume. Pour ce second volume, traitant de la synthèse statistique optimale des systèmes informationnels, nous avons également cru bon de donner un chapitre d'introduction dans lequel est présentée une vue d'ensemble du problème.

L'auteur sera très reconnaissant au lecteur pour toute remarque ou suggestion sur le contenu ou la présentation de la matière dans cette édition française de l'ouvrage. Il tient également à remercier Anne Sokova et les collaborateurs de la rédaction française des Editions Mir du soin apporté au difficile travail de la traduction et de la préparation de cette édition des « Fondements théoriques de la radiotechnique statistique ».

B. Lévine

EXTRAIT DE LA PRÉFACE À L'ÉDITION RUSSE DU SECOND VOLUME

Le plan de cette monographie en deux volumes sur les fondements théoriques de la radiotechnique statistique dont la réalisation s'achève par l'édition de ce second volume, quoique conçu il y a près de cinq ans, n'a pas subi de profonds remaniements. C'est pourquoi d'intéressants résultats obtenus par les mathématiciens et les radioélectriciens ces toutes dernières années n'ont pas pu être suffisamment éclairés ni dans le premier volume, consacré aux méthodes d'analyse, ni dans le second, exposant les méthodes de synthèse. Le cadre limité de ce volume et les délais imposés pour sa mise sous presse nous ont obligés soit d'omettre complètement, soit de nous limiter à un exposé très succinct des branches actuelles, quoiqu'encore insuffisamment élaborées, comme les méthodes non paramétriques, les systèmes adaptatifs et les équations différentielles stochastiques. Nous n'avons pas également étudié les méthodes modernes de recherche des extréma des fonctionnelles se rapportant déjà à la programmation mathématique. Nous nous sommes attachés essentiellement à éclairer la théorie paramétrique de la synthèse. Cette branche de la théorie statistique de la synthèse s'est extrêmement développée et peut déjà dans une certaine mesure être jugée classique.

Ce volume est la seconde partie du cours professé par l'auteur depuis de nombreuses années aux étudiants préparant leur thèse. Bien que ce volume soit destiné tout comme le premier aux étudiants du 3e cycle, aux scientifiques, aux ingénieurs-chercheurs hautement qualifiés et aux mathématiciens travaillant dans l'industrie, nous estimons qu'il permettra également d'attirer de nouveaux adeptes à la radiotechnique statistique.

INTRODUCTION

FORMULATION GÉNÉRALE DU PROBLÈME DE LA SYNTHÈSE OPTIMALE

Dans de nombreuses recherches scientifiques et applications techniques on rencontre un problème qu'on peut formuler sous la forme générale comme suit. Soient les résultats des observations (expériences, essais) fixés soit par un dispositif analogique sous la forme de la réalisation $x(t)$ de durée finie d'un certain processus aléatoire, soit par un dispositif numérique sous la forme d'un ensemble fini de nombres $x_1, \ldots, x_n$ qui sont des valeurs de certaines variables aléatoires. Supposons que la réalisation $x(t)$ soit échantillonnée aux instants $t_1, \ldots, t_n$ appartenant à l'intervalle d'observation et que l'on obtienne ainsi un ensemble fini de nombres $x_i = x(t_i)$, $i = 1, \ldots, n$. Avec cette hypothèse supplémentaire les résultats des observations se présentent toujours sous la forme d'un vecteur de dimension finie

$$\mathbf{x} = (x_1, \ldots, x_n),$$

c'est-à-dire par un point d'un certain sous-ensemble $\mathbf{X}$ de l'espace euclidien (*espace des observations*). La mesure probabiliste sur le sous-ensemble $\mathbf{X}$ est une densité de probabilité multidimensionnelle $W(\mathbf{x})$, $\mathbf{x} \in \mathbf{X}$ définie d'une manière incomplète. On peut, par exemple savoir seulement que $W(\mathbf{x})$ appartient à la classe des fonctions continues (cas non paramétrique). Presque partout dans ce volume nous allons supposer que la densité de probabilité, donnée dans l'espace des observations, appartienne à la classe des fonctions $W(\mathbf{x} \mid \vartheta)$ pour laquelle on connaît la forme de la relation fonctionnelle, mais sur le paramètre vectoriel ϑ on sait seulement qu'il est l'un des éléments du sous-ensemble Θ de l'espace euclidien des paramètres (*cas paramétrique*).

Dans le cas paramétrique le chercheur désire de connaître la vraie valeur du paramètre inconnu ϑ, l'information sur lequel se trouve contenue dans les résultats de l'observation $\mathbf{x}$. Le problème de la synthèse optimale se ramène à la recherche de l'algorithme de traitement (transformation) des résultats des observations, conduisant à un choix optimal (en un certain sens) d'une décision relative à la valeur du paramètre inconnu. Toute décision établit une correspondance entre le point $\mathbf{x}$ de l'espace des observations $\mathbf{X}$ et le point $\gamma(\mathbf{x})$ de l'*espace des décisions* Γ. Chaque décision $\gamma(\mathbf{x})$, en tant que

fonction de l'échantillon observé ou fonctionnelle de la réalisation observée, est appelée *statistique*.

Pour concrétiser la notion de décision optimale nous allons introduire la *fonction de pertes* $\Pi(\gamma, \vartheta)$ tenant compte des conséquences du choix de la décision, et le *critère de qualité* qui est la valeur moyenne de la fonction de pertes sur l'espace des observations, c'est-à-dire la grandeur

$$r(\gamma, \vartheta) = \int_{\mathbf{X}} \Pi[\gamma(\mathbf{x}), \vartheta] W(\mathbf{x} \mid \vartheta) d\mathbf{x},$$

appelée *risque conditionnel*. La décision optimale $\gamma^*(\mathbf{x})$ correspond au minimum du risque conditionnel pour une classe donnée de fonctions $\gamma(\mathbf{x})$ pour ϑ fixé. Autrement dit, pour une suite assez longue d'observations, la décision optimale $\gamma^*(\mathbf{x})$ assure les pertes moyennes minimales par rapport à celles qui apparaissent lorsque l'on adopte toute autre décision $\gamma(\mathbf{x})$.

On obtient un critère de qualité quelque peu différent en changeant de point de vue sur la nature du paramètre inconnu ϑ. Au cours des observations on suppose ce paramètre *aléatoire* dont la densité de probabilité a priori $w(\vartheta)$ donnée sur l'espace Θ soit entièrement connue. On trouve alors la décision optimale en minimisant le risque moyen, c'est-à-dire la moyenne du risque conditionnel sur l'espace des paramètres

$$R(\gamma) = \int_{\Theta} r(\gamma, \vartheta) w(\vartheta) d\vartheta = \int_{\Theta} \int_{\mathbf{X}} \Pi[\gamma(\mathbf{x}), \vartheta] W(\mathbf{x} \mid \vartheta) d\mathbf{x} \, d\vartheta.$$

La décision obtenue ainsi est appelée *décision de Bayes* et la valeur minimale $R_{\min}$, risque de Bayes. Une partie importante de ce volume est consacrée à la synthèse des algorithmes, optimaux dans le sens du critère de qualité de Bayes (minimum du risque moyen).

Souvent on peut obtenir le risque de Bayes même si pour adopter une décision γ on utilise non pas chaque élément de l'échantillon $\mathbf{x} = (x_1, \ldots, x_n)$, mais un certain ensemble de statistiques $g_1(\mathbf{x}), \ldots, g_r(\mathbf{x})$ dont le nombre r est *inférieur* à la taille n de l'échantillon observé. Ceci signifie que la fixation de chacun des éléments de l'échantillon fournit une information superflue du point de vue de la décision optimale et que pour cette décision il suffit d'avoir au lieu du vecteur $\mathbf{x}$ à n dimensions, un vecteur dit des statistiques suffisantes de dimension $r < n$. Il est intéressant de trouver le vecteur des statistiques suffisantes de la dimension r la plus petite possible pour laquelle la propriété de suffisance se conserve encore.

Cette courte introduction suffit pour caractériser le problème de la synthèse optimale. Vu le caractère purement appliqué de cet ouvrage, nous étudions ensuite en détail les méthodes mathématiques de solution de ce problème, séparément pour chacune des deux sortes

de déductions statistiques (décisions): la *vérification des hypothèses statistiques* (premier chapitre) et les *estimations des paramètres* (second chapitre).

La vérification des hypothèses statistiques revient au choix d'une décision à la base des résultats des observations, notamment, adopter une hypothèse (d'un certain ensemble d'hypothèses) sur les états possibles de l'effet étudié. A chaque état on peut faire correspondre un certain nombre s_k, $k = 1, \ldots, m$, l'ensemble de ces nombres sera alors un cas particulier de l'espace des paramètres avec un nombre fini de points. Considérés comme événements aléatoires, ces états forment un groupe complet d'événements. La stratégie attribue à chacun des résultats des observations $\mathbf{x}$ l'une des décisions γ_i, $i = 1, \ldots, m$ selon laquelle l'état s_i est vrai

$$\gamma(\mathbf{x}) = \gamma_i, \quad \mathbf{x} \in \mathbf{X}_i, \quad i = 1, \ldots, m,$$

où $\mathbf{X}_1, \ldots, \mathbf{X}_m$ sont des sous-ensembles disjoints de l'espace des observations. La fonction de pertes est donnée par une matrice carrée $m \times m$ dont les éléments

$$\Pi_{jk} = \Pi(s_j, \gamma_k)$$

sont les coûts de la décision erronée γ_k, lorsque l'état vrai est s_j. Le critère de qualité de la stratégie est le risque conditionnel (somme pondérée des coûts pour une colonne donnée de la matrice, compte tenu de la fréquence d'apparition de chacun d'eux dans une longue suite d'observations) ou bien le risque moyen (la moyenne du risque conditionnel correspondant à la distribution a priori discrète des états).

Dans les problèmes relatifs aux estimations des paramètres l'espace des paramètres est un continuum. L'espace des décisions Γ est, par conséquent, également un continuum, et chaque point de cet espace représente une certaine estimation $\hat{\vartheta}$ du paramètre inconnu ϑ

$$\vartheta = \gamma(\mathbf{x}).$$

Chacune des méthodes mentionnées d'inférence statistique est basée sur la représentation des résultats des observations par un vecteur $\mathbf{x}$ de dimension finie. Les décisions $\gamma(\mathbf{x})$, en tant que fonctions de ce vecteur appelées statistiques, sont des événements aléatoires (pour la vérification des hypothèses) ou des variables aléatoires (lors de l'estimation des paramètres). En ce sens la vérification des hypothèses statistiques peut être appelée *statistique des événements aléatoires*, et l'estimation des paramètres, *statistique des variables aléatoires*.

L'échantillonnage des observations peut parfois conduire à une perte partielle d'information. C'est pourquoi il est très intéressant d'étudier les décisions sous la forme de fonctionnelles $\gamma[x(t)]$ d'une

réalisation du processus aléatoire, qui sont des *statistiques du processus aléatoire*. Les chapitres trois et quatre sont consacrés à cette question.

Les trois derniers chapitres illustrent abondamment les applications des méthodes de synthèse statistique optimale aux problèmes de la détection et de la discrimination des signaux noyés dans les bruits (vérification des hypothèses statistiques), aux problèmes d'extraction des signaux noyés dans les bruits (estimation des paramètres) ainsi qu'aux problèmes de la classification des ensembles (reconnaissance des images).

Chapitre premier

VÉRIFICATION DES HYPOTHÈSES STATISTIQUES

1.1. CHOIX D'UNE DÉCISION

1.1.1. Données de base et formulation du problème. Soit $s_0, \ldots, s_m$ l'ensemble d'*états possibles* (effets de la nature, causes donnant lieu à l'apparition des événements, etc.) représentant un groupe complet et soient $p_0, \ldots, p_m$ $(\sum_{k=0}^{m} p_k = 1)$ les probabilités a priori de ces états. Considérons l'ensemble des résultats des observations $x_1, \ldots, x_n$ (valeurs échantillonnées), fonction de l'état réel, et soit $W_n (x_1, \ldots, x_n \mid s_k)$ la répartition conditionnelle des valeurs échantillonnées, correspondant à l'état s_k, $k = 0, 1, \ldots, m$.

Soient : un *ensemble des décisions* $\gamma_0, \ldots, \gamma_m$ relatives à l'autenticité des états, les *stratégies* $\delta (\gamma_i \mid x_1, \ldots, x_n)$ attribuant à chacun des résultats éventuels des observations $x_1, \ldots, x_n$ l'une des décisions γ_i, $i = 0, 1, \ldots, m$, ainsi que la *fonction de pertes* $\Pi (s_j, \gamma_k)$ tenant compte des conséquences de la stratégie, et enfin le *critère de qualité* f $\{\Pi\}$ de la stratégie, lié à la fonction de pertes.

Le problème que nous allons examiner en détail se trouve à la base de nombreux problèmes techniques et scientifiques. Il peut se formuler comme suit : connaissant les répartitions p_k, $W_n (x_1, \ldots$ $\ldots, x_n \mid s_k)$, l'ensemble des décisions γ_i, la fonction de pertes Π et le critère de qualité f, trouver la *stratégie* δ d'utilisation des résultats des observations $x_1, \ldots, x_n$, la meilleure vis-à-vis du critère adopté. Cette stratégie est une variante de la déduction statistique obtenue d'après les résultats des observations sur les aspects inconnus du phénomène étudié, plus exactement sur le modèle mathématique adopté, lorsque l'information dont nous disposons sur les caractéristiques de ce modèle est incomplète.

Avant d'étudier plus en détail les notions introduites et de formuler le problème du choix d'une décision nous allons donner un exemple très simple.

1.1.2. Transmission des communications binaires en présence des bruits. Soit un système se composant d'une source pouvant éventuellement transmettre deux communications « oui » et « non » transformées respectivement en signaux « 1 » et « 0 », d'une ligne de transmission dont les bruits peuvent transformer à tort un signal « 1 » en signal « 0 » et « 0 » en « 1 » et d'un récepteur reproduisant

les signaux déformés ou non. L'opérateur observant le signal à la sortie du récepteur doit pouvoir déchiffrer le message, c'est-à-dire déterminer d'après le signal reçu la communication transmise.

Si les signaux transmis n'étaient pas soumis à des déformations, le signal reçu permettrait de dire sans erreur quelle communication a été envoyée. Par exemple, le signal « 1 » correspondrait toujours à la communication « oui », autrement dit, la probabilité a posteriori de la communication « oui » pour la réception du signal « 1 » serait alors égale à l'unité. Mais vu les déformations dues aux bruits du signal reçu, on ne peut dire en toute certitude quelle communication a réellement été transmise, c'est-à-dire qu'il peut se faire qu'on reçoit le signal « 1 », alors que c'est la communication « non » qui a été transmise, et le signal « 0 » pour la communication « oui ». Il faut donc donner à l'opérateur à l'avance une règle lui permettant dans cette situation indéterminée d'agir en conséquence, sans se baser sur son intuition ou une appréciation subjective.

Cet exemple, bien que très simple, contient tous les éléments du problème posé. Les communications « non » et « oui » sont deux états incompatibles s_0 et s_1 de la source. Les probabilités a priori de ces états $p_0 = P\{\text{« non »}\}$, $p_1 = P\{\text{« oui »}\} = 1 - p_0$ déterminent la structure statistique de la source des communications, c'est-à-dire indiquent le pourcentage des communications « non » et « oui » dans un message. Les signaux « 1 » et « 0 » sont les résultats possibles des observations. Les probabilités conditionnelles de ces signaux $P\{0 \mid s_0\} = 1 - P\{1 \mid s_0\}$, $P\{1 \mid s_1\} = 1 - P\{0 \mid s_1\}$ sont déterminées par les propriétés probabilistes des bruits de la ligne.

Les grandeurs $P\{0 \mid s_0\}$, $P\{1 \mid s_1\}$ sont les probabilités pour que les signaux « 0 » et « 1 » ne soient pas déformés par les bruits, et $P\{1 \mid s_0\}$, $P\{0 \mid s_1\}$, les probabilités pour que « 0 » devienne « 1 » et « 1 » devienne « 0 ». Dans ce cas l'ensemble des décisions se compose de γ_0 et γ_1, décisions sur la transmission des communications « non » et « oui » respectivement, et la stratégie indique à l'opérateur la décision à prendre lorsqu'il observe le signal « 1 » ou le signal « 0 ».

Dans l'exemple considéré la fonction de pertes doit tenir compte des conséquences qu'entraînent les décisions erronées de l'opérateur et établir le « coût » $\Pi_{01} > 0$ d'une erreur de première espèce (un « non » authentique est interprété comme un «oui») et le « coût » $\Pi_{10} > 0$ pour une erreur de seconde espèce (un « oui » réellement transmis est interprété comme un « non »). En plus des pertes Π_{01}, Π_{10} on aurait pu introduire les grandeurs des gains Π_{00}, Π_{11} (pertes négatives), acquisitions dues aux décisions correctes, ou des dépenses pour des décisions correctes (à condition d'avoir $\Pi_{00} < \Pi_{01}$ et $\Pi_{11} < \Pi_{10}$). Cependant parfois on se limite à l'introduction des pertes liées aux décisions erronées, supposant que les décisions correctes n'entraînent pas de pertes.

En qualité de critère guidant le choix d'une décision on peut utiliser la valeur moyenne des pertes liées aux décisions erronées, pondérée par les probabilités de leur apparition. Ainsi, conformément à ce critère, des deux stratégies possibles on choisit celle pour laquelle la valeur moyenne des pertes est la plus petite. Calculons les pertes moyennes pour les deux stratégies.

L'une des stratégies, δ_0, peut être formulée de la manière suivante : en observant le signal « 0 » il y a lieu d'adopter la décision γ_0 (par conséquent, en observant le signal « 1 » il y a lieu d'adopter la décision γ_1). Dans ce cas les probabilités des décisions erronées sont :

$$P\,\{\text{erreur de } 1^{\text{re}} \text{ espèce}\} = P\,\{\gamma_1 \mid s_0\} = p_0 P\,\{1 \mid s_0\},$$
$$P\,\{\text{erreur de } 2^{\text{e}} \text{ espèce}\} = P\,\{\gamma_0 \mid s_1\} = p_1 P\,\{0 \mid s_1\}$$

et la valeur moyenne R_0 des pertes est

$$R_0 = \Pi_{01} p_0 P\,\{1 \mid s_0\} + \Pi_{10} p_1 P\,\{0 \mid s_1\}. \tag{1.1}$$

La seconde stratégie, δ_1, peut se formuler comme suit : en observant le signal « 0 » il y a lieu d'adopter la décision γ_1 (par conséquent, en observant le signal « 1 » il y a lieu d'adopter la décision γ_0). Dans ce cas les probabilités des décisions erronées sont

$$P\,\{\text{erreur de } 1^{\text{re}} \text{ espèce}\} = P\,\{\gamma_0 \mid s_0\} = p_0 P\,\{0 \mid s_0\},$$
$$P\,\{\text{erreur de } 2^{\text{e}} \text{ espèce}\} = P\,\{\gamma_1 \mid s_1\} = p_1 P\,\{1 \mid s_1\}$$

et la valeur moyenne R_1 des pertes est

$$R_1 = \Pi_{01} p_0 P\,\{0 \mid s_0\} + \Pi_{10} p_1 P\,\{1 \mid s_1\}. \tag{1.2}$$

Le critère de qualité adopté donne la préférence à la stratégie δ_0 si $R_0 < R_1$, c'est-à-dire lorsque

$$R_0 < \frac{1}{2}(R_0 + R_1). \tag{1.3}$$

Comme $R_0 + R_1 = \Pi_{01} p_0 + \Pi_{10} p_1$, en vertu de (1.1) et (1.3) on arrive à la condition suivante :

$$\Pi_{01} p_0 P\,\{1 \mid s_0\} + \Pi_{10} p_1 P\,\{0 \mid s_1\} < \frac{1}{2}(\Pi_{01} p_0 + \Pi_{10} p_1). \tag{1.4}$$

Celle-ci étant satisfaite, on adopte la stratégie δ_0.

En plus des probabilités conditionnelles d'erreurs données par les caractéristiques probabilistes des bruits dans la ligne, la formule (1.4) contient les probabilités a priori des communications et les valeurs des pertes. Il est parfois difficile de trouver ou de fixer les valeurs de p_0, Π_{01}, Π_{10}. C'est là également le point faible du problème général mentionné au § 1.1.1. Lorsqu'il n'y a aucune raison pour supposer que l'erreur de première espèce est nettement supérieure ou inférieure à l'erreur de seconde espèce, on prend les deux pertes Π_{01} et Π_{10} égales. Dans ce cas les valeurs des pertes moyennes sont

simplement proportionnelles à la probabilité d'erreur d'espèce quelconque. Le critère des pertes moyennes minimales devient alors le critère de fréquence minimale d'erreur. Lorsqu'on ne sait rien sur la structure statistique de la source de communication, il ne reste plus qu'à supposer que les communications « oui » et « non » sont transmises avec des probabilités égales, c'est-à-dire à poser: $p_0 = p_1 = \frac{1}{2}$.

Si $\Pi_{01} = \Pi_{10}$ et $p_0 = p_1 = \frac{1}{2}$, la condition (1.4) devient particulièrement simple, on a

$$P\{1 \mid s_0\} + P\{0 \mid s_1\} < 1$$

ou

$$P\{1 \mid s_0\} < P\{1 \mid s_1\}. \tag{1.5}$$

La condition (1.5) signifie que la probabilité de déformation du signal « 0 » est inférieure à celle de reproduction correcte du signal « 1 » (autrement dit, la probabilité d'apparition d'un signal faux « 1 » serait inférieure à la probabilité de non-altération du signal authentique « 1 »). Les pertes moyennes dans ce cas sont $R_0 < 0{,}5$ alors que pour la stratégie δ_1 sous la condition (1.5) les pertes moyennes sont $R_1 > 0{,}5$ *).

1.1.3. Hypothèses simples et hypothèses composites. Arrêtons-nous sur les notions de base introduites au § 1.1.1 pour le problème général du choix des décisions. Parfois ce problème peut se formuler en termes de la théorie de la *vérification des hypothèses statistiques.* Le choix d'une décision consiste alors à adopter ou à rejeter l'hypothèse concernant les états possibles de l'effet étudié d'après les résultats des observations. Ainsi, par exemple, en observant à la sortie d'un récepteur un signal noyé dans les bruits, on peut le prendre pour un bruit (état s_0) ou pour un mélange du signal et du bruit (état s_1). Le choix d'une décision consiste à adopter ou à rejeter l'hypothèse H_0 selon laquelle l'effet observé ne se rapporte qu'au bruit. L'hypothèse contraire H_1 selon laquelle l'effet observé se rapporte au mélange du signal et du bruit est une alternative.

Une classe d'hypothèses est dite *simple* si elle ne contient qu'une seule hypothèse et *composite* si le nombre d'hypothèses est égal ou supérieur à deux. Dans l'exemple ci-dessus l'hypothèse vérifiée

---

*) Notons qu'il aurait été possible de suivre les stratégies suivantes: δ_3 — adopter toujours la décision γ_0, ou δ_4 — adopter toujours la décision γ_1. Les valeurs des pertes moyennes pour ces décisions sont respectivement $R_3 = p_1\Pi_{10}$ et $R_4 = p_0\Pi_{01}$. Si $\Pi_{01} = \Pi_{10} = 1$ et $p_0 = p_1 = \frac{1}{2}$, chacune de ces stratégies est plus mauvaise que δ_0 car on a $R_3 = R_4 = \frac{1}{2}$, mais meilleure que δ_1 car $R_1 > \frac{1}{2}$.

et l'alternative sont simples. Dans le problème de la détection d'un des signaux d'un ensemble, noyés dans le bruit (états $s_1, \ldots, s_m$), la classe des alternatives est composite. Dans une hypothèse composite le nombre d'états peut être tant fini qu'infini; on parle alors d'ensemble dénombrable (discret) ou d'un continuum d'états. Les états possibles sont souvent représentés par des points dans un espace multidimensionnel. Si l'ensemble des états possibles forme un continuum, la distribution a priori des probabilités des états est caractérisée par la densité de probabilité $w_1(s)$ déterminée sur cet ensemble.

1.1.4. Echantillon. Comme nous l'avons noté dans le premier volume, à toute expérience aléatoire on peut faire correspondre une certaine variable aléatoire ξ dont les valeurs possibles sont les résultats des observations fixés après cette expérience aléatoire. Ainsi, le résultat d'une suite de n expériences aléatoires est représenté par n valeurs éventuelles $x_1, \ldots, x_n$ de la variable aléatoire ξ. Chacune de ces valeurs est dite *valeur échantillonnée* et leur ensemble s'appelle *échantillon.* Le nombre n de valeurs échantillonnées, ou le nombre d'éléments de l'échantillon, est généralement appelé *dimension*, ou *taille*, de l'échantillon. Supposons que la distribution de la variable aléatoire ξ soit caractérisée par la fonction de répartition $F_1(x)$ ou la densité de probabilité $w_1(x)$. On dit alors que l'échantillon $x_1, \ldots, x_n$ a été obtenu à partir de la distribution $F_1(x)$ ou $w_1(x)$. La loi de distribution peut dépendre de l'état s_k de l'effet étudié. Pour souligner le fait que l'échantillon dépend de l'état s_k, on désigne la loi de distribution à partir de laquelle l'échantillon a été obtenu par les symboles $F_1(x \mid s_k)$ ou $w_1(x \mid s_k)$.

Chaque échantillon, c'est-à-dire chaque ensemble de n nombres, correspond à un point de l'espace à n dimensions, ou à un vecteur à n dimensions $X = (x_1, \ldots, x_n)$. Si l'on considère (ne serait-ce qu'en pensée) toutes les suites possibles d'expériences, l'ensemble des échantillons possibles remplit un certain domaine de l'espace à n dimensions mentionné. Ce domaine est appellé *espace des échantillons.* La densité de probabilité obtenue à partir de la distribution $w_1(x \mid s)$ avec les éléments indépendants $x_1, \ldots, x_n$ est égale à

$$W_n(x_1, \ldots, x_n \mid s) = \prod_{k=1}^{n} w_1(x_k \mid s), \tag{1.6}$$

car dans ce cas la densité de probabilité conjointe des valeurs échantillonnées est égale au produit des densités de probabilité des éléments de l'échantillon.

Pour un échantillon pris dans la distribution discrète $p_{j|s} = P\{x = x_j \mid s\}$, $j = 1, 2, \ldots, n$, on a

$$P\{x = x_1, \ldots, x = x_n \mid s\} = \prod_{k=1}^{n} p_{k\mid s}. \tag{1.6'}$$

La distribution conjointe des valeurs échantillonnées est appelée *fonction de vraisemblance* de l'échantillon. La fonction de vraisemblance définie dans l'espace des échantillons donne la fréquence d'apparition des points dans un certain domaine de cet espace et détermine la densité de probabilité de l'apparition d'un point quelconque.

Ainsi, on comprend un échantillon comme un ensemble de variables aléatoires, ou comme une variable aléatoire multidimensionnelle (vecteur aléatoire), caractérisée par une certaine distribution appelée fonction de vraisemblance *).

Généralisant la notion d'échantillon, on peut supposer que les résultats des observations sont les valeurs non pas d'une seule variable aléatoire, mais d'un ensemble de variables aléatoires indépendantes, liées par la densité de probabilité à n dimensions $W_n (x_1, \ldots, x_n \mid s)$ dont la forme est donnée par l'état s de l'effet étudié.

Jusqu'à présent nous avons supposé que lorsque l'on prenait les valeurs échantillonnées $x_1, \ldots, x_n$, l'état s_k de l'effet étudié ne changeait pas. Dans l'exemple de la détection ceci signifie que l'effet observé est lié soit seulement au bruit, soit au mélange du signal et du bruit. Cependant, il peut se faire que l'état de l'objet étudié change une ou plusieurs fois durant le temps d'extraction des valeurs échantillonnées. Par exemple, un signal noyé dans les bruits apparaît dans les limites de l'intervalle donné d'observation. L'échantillon est alors *hétérogène*, ses différentes parties appartenant à des distributions différentes $w_1 (x \mid s_k)$, $w_1 (x \mid s_l)$, etc.

Enfin, il faut noter que, si le problème est posé comme mentionné ci-dessus, les déductions statistiques sont faites à partir d'un échantillon de dimension *n donnée*. Parfois il est plus commode de ne pas fixer à l'avance le volume des expériences, mais de le trouver durant les observations, en décidant de continuer ou d'arrêter les observations après chacune des expériences, après avoir soumis à l'analyse statistique les données obtenues à cet instant.

1.1.5. Ensemble de décisions et choix d'une décision. L'ensemble des décisions $\gamma_0, \gamma_1, \ldots, \gamma_m$ est une suite d'assertions logiques déclarant vraie l'une des hypothèses relatives aux états $s_0, s_1, \ldots, s_m$ de l'effet étudié. La stratégie δ établit une correspondance entre un ensemble de décisions et les résultats éventuels des observations, c'est-à-dire l'espace des échantillons **). Ceci signifie que l'espace des échantillons G doit être divisé en $m + 1$ domaines $G_0, G_1, \ldots$

*) Notons qu'à l'opposé du premier volume où les variables aléatoires et les arguments des fonctions de répartition correspondantes étaient désignés par des symboles différents (grecs et latins), ici les éléments de l'échantillon et les arguments de la fonction de vraisemblance correspondante sont désignés par des symboles identiques, ce qui ne donne pas lieu à identifier ces grandeurs.

**) Dans les ouvrages de mathématiques la stratégie est souvent appelée *critère*, ou *test*. Nous utiliserons le terme de critère seulement lorsqu'il s'agit d'un critère de qualité de la stratégie.

. . ., G_m disjoints, la stratégie établit alors la correspondance entre les décisions γ_k et les domaines $G_j(x_1, \ldots, x_n)$. Il est important de souligner que la stratégie doit être formulée *avant* l'observation.

Une stratégie peut être *déterministe* (non randomisée), dans ce cas on fait toujours correspondre à un domaine $G_k(x_1, \ldots, x_n)$ une certaine décision γ_k, autrement dit, si l'échantillon observé se trouve dans le domaine G_k on adopte la décision γ_k, c'est-à-dire que l'on affirme l'authenticité de l'hypothèse suivant laquelle l'effet étudié se trouve dans l'état s_k. Une stratégie peut être *randomisée*, dans ce cas pour des valeurs échantillonnées données $x_1, \ldots, x_n$ on peut adopter une des décisions possibles conformément à une certaine distribution des probabilités. Cette distribution $P\{\gamma_k \mid (x_1, \ldots$ $\ldots, x_n) \in G_j\}$ donne les probabilités conditionnelles des décisions pour un échantillon donné $(\sum_{k=0}^{m} P\{\gamma_k \mid (x_1, \ldots, x_n) \in G_j\} = 1, j =$ $= 0, 1, \ldots, m)$. Pour une stratégie déterministe on a $P\{\gamma_k \mid (x_1, \ldots$ $\ldots, x_n) \in G_j\} = 1$ pour un seul $k = j$, cette probabilité étant nulle pour tous les autres $k \neq j$. Dans cet ouvrage nous ne parlerons que des stratégies déterministes. Cette limitation ne porte en aucun cas atteinte aux stratégies randomisées, présentant un grand intérêt et permettant parfois de simplifier l'analyse mathématique (voir, par exemple, [7]).

1.1.6. Fonction de pertes et critère de qualité d'une stratégie. Vu la nature aléatoire des événements observés, l'utilisation d'une stratégie quelconque, établie à l'avance, est indubitablement liée à une éventualité de décision erronée. Il peut se faire que l'échantillon $x_1, \ldots, x_n$ se trouve dans le domaine G_k, ceci donne lieu à la décision γ_k que le phénomène étudié se trouve à l'état s_k, bien qu'en réalité l'échantillon en question soit lié à un autre état s_j, $j \neq k$. Le fait que, dans une suite de décisions, il y en a qui sont correctes et d'autres erronées, est le coût *inévitable* d'une prise de décision basée sur une information incomplète.

Les conséquences des décisions erronées peuvent être très différentes. Pour en tenir compte analytiquement on introduit une fonction de pertes non négative faisant correspondre à chaque décision erronée, c'est-à-dire à chaque combinaison s_j, γ_k, $k \neq j$, une perte $\Pi_{jk} = \Pi(s_j, \gamma_k) > 0$. De plus, on peut introduire les valeurs des gains (pertes négatives) acquis lorsque la décision se trouve être correcte ou les dépenses liées à des décisions correctes $\Pi_{jj} =$ $= \Pi(s_j, \gamma_j) < \Pi_{jk}$, $k \neq j$. Pour un état donné s_j la valeur moyenne des pertes pour une stratégie donnée δ (c'est-à-dire pour un mode de partition de l'espace des échantillons en domaines G_k et d'établissement de leur correspondance à un ensemble de décisions γ_k) dans une suite suffisamment longue d'expériences, dont les résultats sont fixés par des échantillons de dimension n, est à peu près égale à la moyenne (espérance mathématique) de la fonction de pertes dans

l'espace des échantillons:

$$r_j = \sum_{k=0}^{m} \Pi_{jk} P\{\gamma_k \mid s_j\} = \sum_{k=0}^{m} \Pi_{jk} P\{(x_1, \ldots, x_n) \in G_k \mid s_j\}, \quad (1.7)$$

où $P\{(x_1, \ldots, x_n) \in G_k \mid s_j\}$ est la probabilité conditionnelle pour l'échantillon $(x_1, \ldots, x_n)$ de tomber dans le domaine G_k si en réalité c'est l'état s_j qui a lieu. La moyenne conditionnelle r_j pour un état donné s_j est appelée *fonction conditionnelle de risque*.

On aurait pu prendre la fonction conditionnelle de risque pour critère de qualité de la stratégie et considérer que la meilleure stratégie est celle qui rend minimale la grandeur δ pour toutes les valeurs possibles de r_j. Cependant, la propriété d'optimalité de la stratégie dépendrait alors de l'état s_j. Pour un autre état s_k, $k \neq j$, la stratégie minimisant r_j peut ne pas minimiser r_k.

En prenant la moyenne de la fonction de risque sur tous les états possibles s_j on obtient

$$R = \sum_{j=0}^{m} p_j r_j = \sum_{j=0}^{m} \sum_{k=0}^{m} p_j \Pi_{jk} P\{(x_1, \ldots, x_n) \in G_k \mid s_j\}, \quad (1.8)$$

où p_j est la probabilité a priori de l'état s_j.

La moyenne de la fonction de pertes définie par (1.8), dépendant également de la stratégie et de la distribution des états, est appelée *fonction de risque moyenne*. Cette fonction peut être prise comme critère de la qualité de la stratégie. Dans ce cas la stratégie optimale (pour le critère adopté) correspondra à la plus petite valeur de la fonction moyenne de risque, étant donné la fonction de pertes, la distribution des états et les probabilités conditionnelles des échantillons pour des états donnés. La stratégie optimale définit le mode (parmi la multitude possible) de partition de l'espace des échantillons en domaines disjoints G_k, $k = 0, 1, \ldots, m$, qui, utilisé assez longtemps, assure des pertes moyennes minimales (pour les états possibles de l'effet étudié).

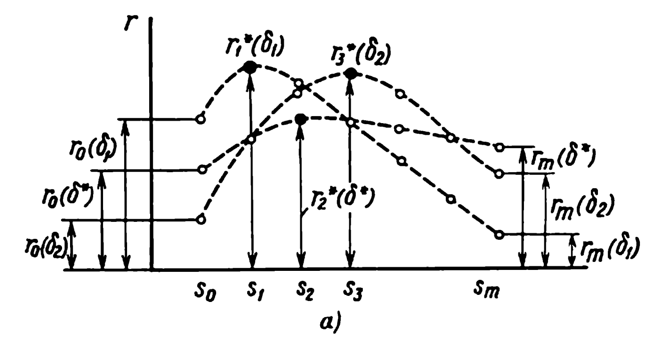
La stratégie optimale, minimisant la fonction moyenne de risque est appelée *stratégie de Bayes*, et la valeur minimale correspondante de la fonction moyenne de risque est appelée *risque de Bayes*.

La théorie exposée a un défaut important. Avant d'utiliser les résultats il faut disposer a priori d'une information assez importante non seulement sur les densités de probabilités conditionnelles des valeurs échantillonnées $W_n(x_1, \ldots, x_n \mid s_j)$, qui peuvent souvent être données avec assez de bien-fondé, mais également sur la fonction de pertes $\Pi(s_j, \gamma_k)$ et sur la distribution a priori des états. Si cette dernière n'est pas connue, pour établir le critère de qualité de la stratégie on ne peut utiliser que la fonction conditionnelle de risque r_j, qui est une fonction de l'argument entier j donnant l'état s_j.

Soient r_{i1}^* la valeur maximale de cette fonction pour la stratégie δ_1 et r_{k2}^* sa valeur maximale pour la stratégie δ_2. On peut consi-

dérer que la meilleure des deux stratégies est celle pour laquelle la valeur maximale de la fonction conditionnelle de risque est la plus petite. Par exemple, si $r^*_{k2} < r^*_{i1}$, la stratégie δ_2 est meilleure que la stratégie δ_1. La stratégie optimale sera donc celle à laquelle correspond un minimum parmi les valeurs maximales des fonctions de risque conditionnelles. Cette stratégie est appelée *stratégie du minimax.*

La stratégie du minimax nous donne la certitude que les pertes moyennes seront inférieures ou égales à une certaine valeur r^*. Bien que dans de nombreux cas cette stratégie soit raisonnable, il



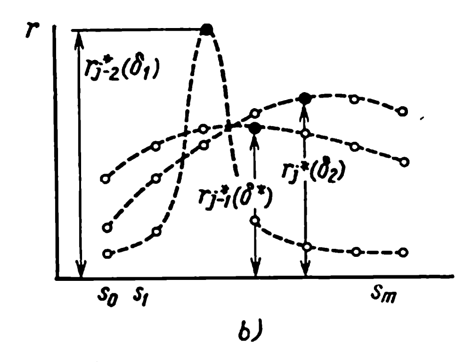


Fig. 1.1. Application de la stratégie du minimax: a) acceptable; b) non acceptable (la stratégie δ_1 est meilleure que δ^*)

peut y avoir des cas où elle s'avère trop prudente. Il est alors préférable d'utiliser une autre stratégie, pour laquelle la valeur maximale du risque conditionnel est supérieure à r^*. Si, par exemple, dans l'un des états la stratégie δ donne un risque conditionnel un peu supérieur à r^*, et dans tous les autres états un risque de beaucoup inférieur à ceux qui correspondent à la stratégie du minimax δ^*, il faut donner la préférence à la stratégie δ. C'est ce que l'on peut voir sur la figure 1.1 où l'on a représenté des cas typiques quand il est

bon d'utiliser la stratégie du minimax (fig. 1.1,*a*) et quand elle se trouve être trop prudente (fig. 1.1,*b*).

On peut montrer (cf., par exemple, [14], page 91) que toute stratégie du minimax est un *cas spécial de la stratégie de Bayes pour la moins favorable des distributions a priori* $(p_j)_{mm}$ des états s_j, $j = 0, \ldots, m$, pour laquelle le risque moyen minimal (de Bayes) a la plus grande valeur parmi celles du risque moyen calculées pour une stratégie de Bayes pour toute autre distribution *). Malheureusement, il n'existe pas de méthode générale permettant de trouver la distribution $(p_j)_{mm}$ a priori la plus défavorable. Cependant, il a été démontré (cf. [14], § 3.5) que la stratégie de Bayes à laquelle correspondent les mêmes risques conditionnels $r_j = r$, $j = 0, 1, \ldots, m$ pour tous les états, est celle du minimax. Ceci peut être utilisé pour trouver la plus défavorable des distributions et, par conséquent, la stratégie du minimax. Mais il ne faut pas oublier que l'égalité des risques conditionnels pour des stratégies *non bayesiennes* ne donne pas la stratégie du minimax. D'un autre côté, si l'égalité des risques conditionnels ne peut être réalisée, ceci ne veut pas encore dire que la stratégie du minimax n'existe pas. Pour trouver cette dernière il y a lieu d'utiliser d'autres méthodes.

Lorsque l'on connaît la distribution a priori des états s_j sans pouvoir se donner les valeurs des pertes Π_{jk}, on peut essayer une méthode quelque peu différente de choix des stratégies. En utilisant la formule de Bayes [cf. (1.19) dans le tome I] nous allons trouver la probabilité a posteriori de l'état s_j, lorsque l'on observe l'échantillon $(x_1, \ldots, x_n)$

$$P\{s_j \mid x_1, \ldots, x_n\} = \frac{p_j W_n(x_1, \ldots, x_n \mid s_j)}{\sum_{k=0}^{m} p_k W_n(x_1, \ldots, x_n \mid s_k)}. \tag{1.9}$$

Les probabilités a posteriori mentionnées donnent la caractéristique la plus complète de l'état s_j de l'effet étudié dans le cas où celle-ci doit être obtenue à partir des valeurs échantillonnées de la variable aléatoire observée. Il se trouve ainsi naturel d'adopter le critère suivant : parmi les hypothèses concernant les états s_j, $j = 0, 1, \ldots, m$, on considère vraie celle pour laquelle la probabilité a posteriori (1.9) est maximale. Ainsi, le critère de qualité de la stratégie est le *maximum de la probabilité a posteriori*.

Ce critère permet de trouver la règle de partition de l'espace des échantillons. Au domaine G_k on fait correspondre les échantillons $(x_1, \ldots, x_n)$ pour lesquels pour tous les $j \neq k$ on a

$$P\{s_k \mid x_1, \ldots, x_n\} \geqslant P\{s_j \mid x_1, \ldots, x_n\}. \tag{1.9'}$$

Si l'on n'a pas de données a priori ni sur la distribution des probabilités des états, ni sur les pertes, on peut utiliser le critère dit

*) Les conditions d'existence de cette distribution la moins favorable ont été formulées dans l'ouvrage [14], § 3.1, théorème 3.14.

du maximum de vraisemblance en vertu duquel, lorsque l'on observe l'échantillon $x_1, \ldots, x_n$, on adopte celle des hypothèses relatives à l'état s_j pour laquelle la fonction de vraisemblance $W_n(x_1, \ldots$ $\ldots, x_n \mid s_j)$ est supérieure aux autres fonctions de vraisemblance $W_n(x_1, \ldots, x_n \mid s_k)$, $k = 0, 1, \ldots, j-1, j+1, \ldots, m$. Ce critère est un cas particulier du critère du maximum de probabilité a posteriori, à condition toutefois que tous les états soient équiprobables, c'est-à-dire que $p_j = \frac{1}{m+1}$.

Un autre critère de qualité de la stratégie se base sur la *quantité moyenne* d'information contenue dans la stratégie adoptée δ sur l'événement étudié s (caractérisé par $m+1$ états $s_0, \ldots, s_m$ et par leurs distributions a priori $p_0, \ldots, p_m$). Par définition de la quantité moyenne d'information [13] on a

$$I(\delta, s) = \sum_{j=0}^{m} \sum_{k=0}^{m} P\{s_j \text{ et } \gamma_k\} \ln \frac{P\{s_j \text{ et } \gamma_k\}}{p_j P\{\gamma_k\}} =$$

$$= H(s) - H(s \mid \delta) = H(s) - \sum_{k=0}^{m} P\{\gamma_k\} H(s \mid \gamma_k), \qquad (1.10)$$

où

$$H(s) = -\sum_{j=0}^{m} p_j \ln p_j \geqslant 0$$

est l'*entropie*, caractérisant l'indétermination a priori des états, et

$$H(s \mid \gamma_k) = -\sum_{j=0}^{m} P\{s_j \mid \gamma_k\} \ln P\{s_j \mid \gamma_k\} \geqslant 0$$

l'*entropie conditionnelle* de l'événement s une fois la décision γ_k prise. De plus,

$$P\{s_j \mid \gamma_k\} = \frac{p_j}{P\{\gamma_k\}} \int \ldots \int_{G_k} W_n(x_1, \ldots, x_n \mid s_j)\, dx_1 \ldots dx_n,$$

$$P\{\gamma_k\} = \sum_{j=0}^{m} p_j \int \ldots \int_{G_k} W_n(x_1, \ldots, x_n \mid s_j)\, dx_1 \ldots dx_n,$$

$$H(s \mid \delta) = \sum_{k=0}^{m} P\{\gamma_k\} H(s \mid \gamma_k), \qquad 0 \leqslant H(s \mid \delta) \leqslant H(s).$$

On considère la meilleure la stratégie δ qui fournit le maximum d'information $I(\delta, s)$, c'est-à-dire minimise la valeur moyenne de l'entropie conditionnelle $H(s \mid \delta)$ (ou, comme on dit dans la théorie de l'information, de la *non-fiabilité*). Autrement dit, la stratégie

satisfaisant à la condition $\max_{\delta} I(\delta, s)$ garantit une perte moyenne minimale d'information, liée à la prise d'une décision à partir des valeurs échantillonnées *).

1.2. VÉRIFICATION D'UNE HYPOTHÈSE SIMPLE CONTRE UNE ALTERNATIVE SIMPLE

1.2.1. Probabilités de décisions correctes et de décisions erronées. Examinons un problème simple de la théorie exposée dans ce chapitre, c'est-à-dire le problème de vérification des hypothèses simples. Soit un certain nombre de valeurs observées $x_1, \ldots, x_n$ (échantillon de taille n). On sait que ces valeurs appartiennent à l'une des distributions $W_n(x_1, \ldots, x_n \mid s_0)$ ou $W_n(x_1, \ldots, x_n \mid s_1)$ caractérisant les états s_0 et s_1 incompatibles de l'effet étudié. Le problème revient à indiquer le meilleur (en un certain sens) algorithme de dépouillement des données observées et à décider à laquelle des distributions mentionnées appartient l'échantillon obtenu.

Désignons par H_0 et H_1 les hypothèses selon lesquelles les valeurs échantillonnées appartiennent aux distributions $W_n(x_1, \ldots, x_n \mid s_0)$ et $W_n(x_1, \ldots, x_n \mid s_1)$ respectivement, et par γ_0 et γ_1 les décisions consistant à adopter ou à rejeter l'hypothèse H_0. L'hypothèse H_1 est une alternative simple de H_0, on peut donc prendre en considération seulement l'hypothèse H_0. Il est évident qu'en rejetant l'hypothèse H_0 on adopte l'hypothèse H_1. Pour les procédures *non randomisées* considérées ici de vérification des hypothèses (voir § 1.1.5) le problème revient à établir avant observation la stratégie selon laquelle à chaque échantillon $x_1, \ldots, x_n$ on fait correspondre l'une des décisions γ_0 où γ_1, autrement dit, à formuler la stratégie permettant d'adopter ou de rejeter l'hypothèse H_0 à partir de données accumulées lors de l'observation de l'effet étudié. Etablir cette loi équivaut à diviser l'espace à n dimensions des échantillons $(x_1, \ldots, x_n)$ en deux domaines G_0 et G_1 disjoints. Si l'échantillon considéré tombe dans le domaine G_0, on adopte l'hypothèse H_0, et s'il tombe dans le domaine G_1, on la rejette (c'est-à-dire que l'on adopte l'hypothèse H_1). Ainsi

$$(x_1, \ldots, x_n) \in G_0 \rightarrow \gamma_0,$$

$$(x_1, \ldots, x_n) \in G_1 \rightarrow \gamma_1,$$

où le symbole $\in$ désigne l'appartenance d'un point à un domaine donné de l'espace.

Le domaine G_0 d'acceptation de l'hypothèse est dit domaine *admissible*, et le domaine G_1 de rejet de l'hypothèse est appelé domai-

*) Un critère informationnel plus complet de la qualité est lié à la maximisation de la *valeur de l'information*, si toutefois la quantité d'information obtenue ne dépasse pas une certaine limite [2].

ne *critique*. L'équation de la surface $D(x_1, \ldots, x_n) = \text{const}$ délimitant ces deux domaines dans l'espace à n dimensions est l'expression analytique de la stratégie.

Quelle que soit la stratégie établie à l'avance, il est inévitable qu'à côté des décisions correctes on prenne également (vu la nature aléatoire de l'échantillon) des décisions *erronées*. Deux types d'erreurs sont possibles. L'*erreur de première espèce* apparaît lorsque l'échantillon tombe dans le domaine critique G_1 alors qu'en réalité c'est l'état s_0 qui a lieu. Par là même l'hypothèse H_0 se trouve rejetée, bien qu'en réalité elle soit vraie. L'*erreur de seconde espèce* apparaît lorsque l'échantillon se trouve dans le domaine G_0 alors qu'en réalité c'est l'état s_1 qui a lieu. On adopte alors une hypothèse fausse. D'une manière analogue on peut considérer deux types de décisions correctes: une hypothèse vraie est adoptée (l'échantillon se trouve dans le domaine G_0 lorsque c'est l'état s_0 qui a lieu) et une hypothèse fausse est rejetée (l'échantillon se trouve dans le domaine G_1 lorsque c'est l'état s_1 qui a lieu).

On peut facilement écrire les expressions des probabilités conditionnelles des erreurs pour un état donné de l'effet étudié. La probabilité conditionnelle α de l'erreur de première espèce est

$$\alpha = P\{\gamma_1 \mid H_0\} = P\{(x_1, \ldots, x_n) \in G_1 \mid s_0\} = \\ = \int \underset{G_1}{\ldots} \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n. \qquad (1.11)$$

La probabilité conditionnelle d'une décision correcte d'adopter l'hypothèse vraie H_0 complète la probabilité indiquée jusqu'à l'unité, c'est-à-dire que l'on a

$$P\{\gamma_0 \mid H_0\} = P\{(x_1, \ldots, x_n) \in G_0 \mid s_0\} = \\ = \int \underset{G_0}{\ldots} \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n = \\ = 1 - \int \underset{G_1}{\ldots} \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n = 1 - \alpha. \qquad (1.11')$$

La probabilité conditionnelle β de l'erreur de seconde espèce est

$$\beta = P\{\gamma_0 \mid H_1\} = P\{(x_1, \ldots, x_n) \in G_0 \mid s_1\} = \\ = \int \underset{G_0}{\ldots} \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n. \qquad (1.12)$$

La probabilité conditionnelle d'une décision correcte de rejeter l'hypothèse fausse H_1 est égale à la différence entre l'unité et β,

en effet

$$P\{\gamma_1 \mid H_1\} = P\{(x_1, \ldots, x_n) \in G_1 \mid s_1\} =$$

$$= \int_{G_1} \ldots \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n =$$

$$= 1 - \int_{G_0} \ldots \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n = 1 - \beta. \quad (1.12')$$

La probabilité α de l'erreur de première espèce (c'est-à-dire la probabilité de rejeter l'hypothèse vraie H_0) est parfois appelée *niveau de signification*, et la probabilité $1 - \beta$ de rejeter une hypothèse fausse est appelée *puissance* de la stratégie ou du critère *).

Sachant que les probabilités a priori des états s_0 et s_1 sont q et $p = 1 - q$ respectivement, on peut à l'aide des formules (1.11) à (1.12') trouver les probabilités a priori d'adopter les décisions γ_0 et γ_1 :

$$P\{\gamma_0\} = qP\{\gamma_0 \mid H_0\} + pP\{\gamma_0 \mid H_1\} = q(1 - \alpha) + p\beta, \quad (1.13)$$

$$P\{\gamma_1\} = pP\{\gamma_1 \mid H_1\} + qP\{\gamma_1 \mid H_0\} = p(1 - \beta) + q\alpha, \quad (1.14)$$

donnant les fréquences d'apparition de différentes décisions dans une longue suite de décisions. Dans les formules (1.13) et (1.14) les premiers termes sont les probabilités a priori de décisions correctes, et les seconds, les probabilités a priori des erreurs.

Pour un échantillon de taille donnée les probabilités d'erreurs de première et de seconde espèce ne peuvent être simultanément aussi petites que l'on veut. Par exemple, pour abaisser le niveau de signification il faut diminuer le domaine critique G_1. Le domaine admissible augmentera évidemment et la sensibilité de la stratégie sera moindre vis-à-vis des erreurs de seconde espèce. Ainsi, pour formuler l'une ou l'autre des stratégies il faut partir des considérations raisonnables. On utilise pour cela les critères de qualité dont certains ont été mentionnés au § 1.1.6. Nous allons appliquer ces critères au problème de la vérification des hypothèses.

---

*) Parfois on introduit dans l'espace des échantillons la fonction de décision $\Phi(X)$ qui pour une stratégie non randomisée est égale à

$$\Phi(X) = \begin{cases} 1, & X \in G_1, \\ 0, & X \in G_0, \end{cases}$$

c'est-à-dire sert en un certain sens de compteur des cas où l'échantillon $X = (x_1, \ldots, x_n)$ tombe dans le domaine critique. A l'aide de la fonction de décision $\Phi(X)$ les formules (1.11) et (1.12') peuvent être écrites comme les moyennes conditionnelles

$$\alpha = m_1\{\Phi(X) \mid H_0\}, \quad 1 - \beta = m_1\{\Phi(X) \mid H_1\}.$$

Pour une stratégie randomisée, $\Phi(X)$ est une certaine fonction de répartition.

1.2.2. Stratégie de Bayes. Introduisons tout d'abord la fonction de pertes attribuant un coût Π_{jk}, $j=0, 1$, $k=0, 1$, à chacune des quatre combinaisons γ_0 et H_0, γ_0 et H_1, γ_1 et H_0, γ_1 et H_1. Il est commode de représenter le coût sous la forme d'une matrice de coût :

$$\Pi = \begin{pmatrix} \Pi_{00} & \Pi_{01} \\ \Pi_{10} & \Pi_{11} \end{pmatrix}, \quad \Pi_{01} > \Pi_{00}, \quad \Pi_{10} > \Pi_{11}, \tag{1.15}$$

dont les lignes correspondent aux hypothèses H_0 et H_1, et les colonnes aux décisions γ_0 et γ_1. Suivant la diagonale principale on a les coûts des décisions correctes *), et suivant l'autre diagonale, les coûts (pertes) liés aux décisions erronées. La valeur moyenne des pertes, pondérée par les probabilités de leur apparition (ou *risque moyen*) est égale à

$$R = qr_0 + pr_1, \tag{1.16}$$

où

$$r_0 = \Pi_{00}P\{\gamma_0 \mid H_0\} + \Pi_{01}P\{\gamma_1 \mid H_0\} = \Pi_{00}(1-\alpha) + \Pi_{01}\alpha, \tag{1.17}$$

$$r_1 = \Pi_{10}P\{\gamma_0 \mid H_1\} + \Pi_{11}P\{\gamma_1 \mid H_1\} = \Pi_{10}\beta + \Pi_{11}(1-\beta) \tag{1.17'}$$

sont les risques conditionnels correspondant aux états s_0 et s_1.

En portant (1.17) et (1.17′) dans (1.16), on obtient après des transformations simples

$$R = q\Pi_{00} + p\Pi_{10} + q(\Pi_{01} - \Pi_{00})\alpha - p(\Pi_{10} - \Pi_{11})(1-\beta). \tag{1.18}$$

Prenons en qualité de critère de la stratégie le minimum du risque moyen R. Ici le risque moyen dépend du domaine critique G_1 par l'intermédiaire des grandeurs α et $1-\beta$. En portant dans (1.18) les expressions de ces grandeurs données par (1.11) et (1.12′) on trouve

$$R = q\Pi_{00} + p\Pi_{10} - \int \ldots \int_{G_1} [p(\Pi_{10} - \Pi_{11})W_n(x_1, \ldots, x_n \mid s_1) - \\ - q(\Pi_{01} - \Pi_{00})W_n(x_1, \ldots, x_n \mid s_0)]\, dx_1 \ldots dx_n. \tag{1.19}$$

Comme $q\Pi_{00} + p\Pi_{10}$ est une constante, on obtiendra le minimum du risque moyen R quand dans (1.19) l'expression sous

*) Si $\Pi_{00} < 0$, $\Pi_{11} < 0$, ces grandeurs peuvent être interprétées comme les gains pour des décisions correctes.

l'intégrale est non négative *), c'est-à-dire quand le domaine critique G_1 de l'espace des échantillons (où l'hypothèse H_0 est rejetée) ne contient que les points pour lesquels on a

$$p\,(\Pi_{10}-\Pi_{11})\,W_n\,(x_1,\,\ldots,\,x_n\mid s_1)\geqslant$$
$$\geqslant q\,(\Pi_{01}-\Pi_{00})\,W_n\,(x_1,\,\ldots,\,x_n\mid s_0)$$

ou

$$\frac{pW_n\,(x_1,\,\ldots,\,x_n\mid s_1)}{qW_n\,(x_1,\,\ldots,\,x_n\mid s_0)}\geqslant\frac{\Pi_{01}-\Pi_{00}}{\Pi_{10}-\Pi_{11}}. \tag{1.20}$$

Ainsi, la stratégie optimale basée sur le critère du minimum de risque moyen, ou *stratégie de Bayes*, peut se formuler de la manière suivante: on adopte la décision γ_1 (rejet de l'hypothèse H_0), si pour l'échantillon observé on a l'inégalité (1.20), et la décision γ_0 (acceptation de l'hypothèse H_0), si on a l'inégalité contraire de (1.20).

L'équation de la surface séparant dans ce cas le domaine critique et le domaine admissible de l'espace des échantillons est

$$D\,(x_1,\,\ldots,\,x_n)=\frac{pW_n\,(x_1,\,\ldots,\,x_n\mid s_1)}{qW_n\,(x_1,\,\ldots,\,x_n\mid s_0)}=\frac{\Pi_{01}-\Pi_{00}}{\Pi_{10}-\Pi_{11}}. \tag{1.21}$$

Le membre $\frac{pW_n(x_1,\,\ldots,\,x_n\mid s_1)}{qW_n(x_1,\,\ldots,\,x_n\mid s_0)}$ de cette équation est appelé *rapport de vraisemblance généralisé* **). La vérification de l'hypothèse simple H_0 consiste à calculer le rapport de vraisemblance généralisé et à le comparer avec un seuil constant c^* qui pour la stratégie de Bayes est égal à ***)

$$c^*=\frac{\Pi_{01}-\Pi_{00}}{\Pi_{10}-\Pi_{11}}. \tag{1.22}$$

En désignant par G_1^* le domaine critique déterminé par l'inégalité (1.20), et par α^* et $1-\beta^*$ le niveau de signification et la puissance correspondant à G_1^*, le minimum de risque moyen R^*,

*) Remarquons que comme pour tout sous-ensemble G_1' de l'ensemble G_1 ($G_1'\subset G_1$) on a l'inégalité

$$\int\limits_{G_1'}\cdots\int f\,(x_1,\,\ldots,\,x_n)\,dx_1\,\ldots\,dx_n\leqslant\int\limits_{G_1}\cdots\int f\,(x_1,\,\ldots,\,x_n)\,dx_1\,\ldots\,dx_n$$

pour $f\,(x_1,\,\ldots,\,x_n)\geqslant 0$, l'intégrale dans (1.19) est maximale si et seulement si le domaine d'intégration G_1 contient tous les points de l'espace des échantillons pour lesquels la fonction sous l'intégrale n'est pas négative.

**) A la différence de la grandeur $l\,(x_1,\,\ldots,\,x_n)=\frac{W_n(x_1,\,\ldots,\,x_n\mid s_1)}{W_n(x_1,\,\ldots,\,x_n\mid s_0)}$ appelée simplement *rapport de vraisemblance*. (Certains auteurs préfèrent le terme de coefficient de vraisemblance.)

***) Bien que lorsque l'on a trouvé la règle optimale mentionnée on n'avait considéré que des règles déterministes, le résultat obtenu reste vrai pour le cas où le risque moyen est minimisé pour une classe plus étendue comprenant également les règles randomisées (cf., par exemple, [9, § 19.1.2.] ou [10, § 2.3])

ou le *risque de Bayes*, s'écrit en vertu de (1.18) sous la forme:

$$R^* = q\Pi_{00} + p\Pi_{10} + q(\Pi_{01} - \Pi_{00})\alpha^* - \\ - p(\Pi_{10} - \Pi_{11})(1 - \beta^*). \quad (1.23)$$

En déterminant les probabilités conditionnelles des erreurs α^* et β^* on peut éviter les difficultés liées aux calculs des intégrales multiples (1.11) et (1.12) et ramener le problème au calcul d'une intégrale simple. La fonction

$$l(x_1, \ldots, x_n) = \frac{W_n(x_1, \ldots, x_n \mid s_1)}{W_n(x_1, \ldots, x_n \mid s_0)}, \quad (1.24)$$

c'est-à-dire le *rapport de vraisemblance* est une grandeur aléatoire non négative obtenue par la transformation fonctionnelle *) de n grandeurs aléatoires $x_1, \ldots, x_n$. Désignons par $W_{10}(y)$ la densité de probabilité du rapport de vraisemblance à condition que l'échantillon soit issu de la distribution $W_n(x_1, \ldots, x_n \mid s_0)$ et par $W_{11}(y)$ celle du rapport de vraisemblance à condition que l'échantillon soit issu de la distribution $W_n(x_1, \ldots, x_n \mid s_1)$. A partir de (1.11) et (1.12) et de la définition du domaine critique G_1^* [cf. (1.20) et (1.22)] on a

$$\alpha^* = P\{(x_1, \ldots, x_n) \in G_1^* \mid s_0\} = \\ = P\{l(x_1, \ldots, x_n) \geqslant \mu c^* \mid s_0\} = \int_{\mu c^*}^{\infty} W_{10}(y)\,dy = 1 - F_{10}(\mu c^*), \quad (1.25)$$

$$\beta^* = P\{(x_1, \ldots, x_n) \in G_0^* \mid s_1\} = \\ = P\{l(x_1, \ldots, x_n) < \mu c^* \mid s_1\} = F_{11}(\mu c^*), \quad (1.26)$$

où F_{10}, F_{11} sont les fonctions de répartition du rapport de vraisemblance pour les hypothèses H_0 et H_1 respectivement et $\mu = \frac{q}{p}$.

1.2.3. Maximum de probabilité a posteriori et maximum de vraisemblance. Calculons les probabilités a posteriori pour l'effet étudié de se trouver dans l'état s_0 ou dans l'état s_1 d'après l'observation de l'échantillon $x_1, \ldots, x_n$. On a à partir de (1.9)

$$P\{s_0 \mid x_1, \ldots, x_n\} = \frac{qW_n(x_1, \ldots, x_n \mid s_0)}{qW_n(x_1, \ldots, x_n \mid s_0) + pW_n(x_1, \ldots, x_n \mid s_1)}, \quad (1.27)$$

$$P\{s_1 \mid x_1, \ldots, x_n\} = \frac{pW_n(x_1, \ldots, x_n \mid s_1)}{qW_n(x_1, \ldots, x_n \mid s_0) + pW_n(x_1, \ldots, x_n \mid s_1)}. \quad (1.28)$$

Adoptons la stratégie suivante: étant donné l'échantillon $x_1, \ldots, x_n$ on accepte l'hypothèse H_0 si $P\{s_0 \mid x_1, \ldots, x_n\} >$

*) La transformation fonctionnelle $z = l(x_1, \ldots, x_n)$ donne l'image des points de l'espace à n dimensions des échantillons sur l'axe réel (domaine unidimensionnel).

$> P\{s_1 \mid x_1, \ldots, x_n\}$ (décision γ_0), et on rejette cette hypothèse si $P\{s_0 \mid x_1, \ldots, x_n\} \leqslant P\{s_1 \mid x_1, \ldots, x_n\}$ (décision γ_1). En utilisant (1.27) et (1.28) on peut formuler cette stratégie comme suit: on adopte la décision γ_1 (rejet de l'hypothèse H_0) si pour l'échantillon observé on a l'inégalité

$$\frac{pW_n(x_1, \ldots, x_n \mid s_1)}{qW_n(x_1, \ldots, x_n \mid s_0)} \geqslant 1, \tag{1.29}$$

et on adopte la décision γ_0 (acceptation de l'hypothèse H_0) pour l'inégalité contraire de (1.29).

Ainsi, il correspond au maximum de probabilité a posteriori un domaine critique de l'espace des échantillons dont les points satisfont à l'inégalité (1.29). La procédure de vérification de l'hypothèse simple H_0 se réduit dans ce cas au calcul du rapport de vraisemblance généralisé et à sa comparaison à l'unité. En comparant (1.29) et (1.20), il est facile de remarquer que la stratégie envisagée est simplement un cas particulier de la décision de Bayes pour le seuil $c^* = 1$ [cf. (1.22)]. Ceci correspond à des coûts égaux des décisions γ_0 et γ_1 ou à des coûts égaux des erreurs $\Pi_{10} = \Pi_{01} = \Pi$ si l'on pose $\Pi_{00} = \Pi_{11} = 0$. Dans ce dernier cas le risque moyen se trouve être égal à [cf. (1.18)]

$$R = (q\alpha + p\beta)\,\Pi, \tag{1.30}$$

c'est-à-dire est égal à la probabilité a priori de l'erreur de première ou de seconde espèce au facteur constant Π près. Par conséquent, la stratégie basée sur le critère du maximum de probabilité a posteriori *minimise la probabilité d'erreur a priori*. Autrement dit, pour une longue suite de décisions, cette stratégie assure une fréquence maximale de *décisions correctes*.

Si lors de l'élaboration de la stratégie on n'a pas de données concernant les probabilités a priori des états s_0 et s_1, on peut au lieu du critère mentionné utiliser le critère du maximum de vraisemblance conformément auquel, pour l'échantillon $x_1, \ldots, x_n$ observé, on adopte l'hypothèse donnant la plus grande des valeurs de la fonction de vraisemblance de l'échantillon. Ainsi, on accepte l'hypothèse H_0 si

$$W_n(x_1, \ldots, x_n \mid s_0) > W_n(x_1, \ldots, x_n \mid s_1)$$

(décision γ_0), et on rejette cette hypothèse si

$$W_n(x_1, \ldots, x_n \mid s_0) \leqslant W_n(x_1, \ldots, x_n \mid s_1)$$

(décision γ_1). Autrement dit, on adopte la décision γ_1 si pour l'échantillon observé on a

$$l(x_1, \ldots, x_n) = \frac{W_n(x_1, \ldots, x_n \mid s_1)}{W_n(x_1, \ldots, x_n \mid s_0)} \geqslant 1, \tag{1.31}$$

et on adopte la décision γ_0 pour l'inégalité contraire de (1.31). La procédure de vérification d'une hypothèse simple suivant le critère

du maximum de vraisemblance se réduit au calcul du *rapport de vraisemblance* et à la comparaison de ce dernier à l'unité. A la différence de la stratégie correspondant au maximum de probabilité a posteriori, ici le rapport de vraisemblance généralisé

$$\frac{pW_n(x_1, \ldots, x_n \mid s_1)}{qW_n(x_1, \ldots, x_n \mid s_0)}$$

est remplacé par le rapport de vraisemblance

$$\frac{W_n(x_1, \ldots, x_n \mid s_1)}{W_n(x_1, \ldots, x_n \mid s_0)}.$$

Par conséquent, la stratégie basée sur le critère du maximum de vraisemblance est un cas particulier de la stratégie basée sur le critère du maximum de probabilité a posteriori, lorsque les deux états possibles s_0 et s_1 de l'effet étudié sont *équiprobables*, c'est-à-dire lorsque l'on a $p = q = \frac{1}{2}$.

1.2.4. Critère de Neumann-Pearson. Le critère de Neumann-Pearson indique une autre méthode d'élaboration de la stratégie en l'absence d'information a priori sur les pertes et les probabilités des différents états. Conformément à ce critère on choisit la stratégie pour laquelle la probabilité d'erreur de seconde espèce β est aussi petite que possible, à condition que la probabilité d'erreur de première espèce ne soit pas supérieure à un α donné. Autrement dit, la stratégie basée sur le critère de Neumann-Pearson a *la plus grande puissance* parmi les stratégies pour lesquelles le niveau de signification ne dépasse pas α.

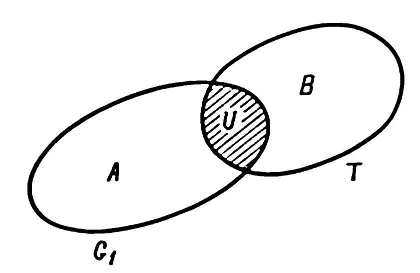


Fig. 1.2. Domaines de l'espace des échantillons

Montrons qu'à cette stratégie correspond le domaine critique G_1 de l'espace des échantillons, contenant seulement les points pour lesquels

$$l(x_1, \ldots, x_n) = \frac{W_n(x_1, \ldots, x_n \mid s_1)}{W_n(x_1, \ldots, x_n \mid s_0)} \geqslant c, \tag{1.32}$$

le seuil c étant choisi à partir de la condition

$$P\{l(x_1, \ldots, x_n) \geqslant c \mid s_0\} = \int_c^\infty W_{10}(y)\, dy = \alpha. \tag{1.33}$$

Soit T un domaine quelconque de l'espace des échantillons satisfaisant à la seule condition

$$P\{(x_1, \ldots, x_n) \in T \mid s_0\} \leqslant \alpha. \qquad (1.34)$$

Désignons par U l'intersection des domaines G_1 et T, par A, la partie du domaine G_1 n'appartenant pas à T et par B, la partie du domaine T n'appartenant pas à G_1 (fig. 1.2) *). Comme les points du domaine A satisfont à l'inégalité (1.32), on a

$$\int_A \cdots \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n \geqslant$$

$$\geqslant c \int_A \cdots \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n.$$

Par suite

$$\int_{G_1} \cdots \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n =$$

$$= \int_A \cdots \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n +$$

$$+ \int_U \cdots \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n \geqslant$$

$$\geqslant c \int_A \cdots \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n +$$

$$+ \int_U \cdots \int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n =$$

$$= c \int_{G_1} \cdots \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n -$$

$$- c \int_U \cdots \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n +$$

$$+ \int_U \cdots \int W_n(x_1, \ldots x_n \mid s_1)\, dx_1 \ldots dx_n. \qquad (1.35)$$

Comme en vertu de (1.33) et (1.34) on a

$$\alpha = \int_{G_1} \cdots \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n \geqslant$$

$$\geqslant \int_T \cdots \int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n,$$

*) Les cas où $A \equiv G_1$, $B \equiv T$, $U \equiv 0$ ne sont pas exclus.

il vient de (1.35) :

$$\int_{G_1}\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n \geqslant$$

$$\geqslant c \int_T\cdots\int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n -$$

$$- c \int_U\cdots\int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n +$$

$$+ \int_U\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n =$$

$$= c \int_B\cdots\int W_n(x_1, \ldots, x_n \mid s_0)\, dx_1 \ldots dx_n +$$

$$+ \int_U\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n .$$

Le domaine B ne contient pas de points appartenant au domaine G_1, donc, en vertu de (1.32), on a pour ce domaine

$$cW_n(x_1, \ldots, x_n \mid s_0) > W_n(x_1, \ldots, x_n \mid s_1).$$

Par conséquent

$$\int_{G_1}\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n >$$

$$> \int_B\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n +$$

$$+ \int_U\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n =$$

$$= \int_T\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n, \qquad (1.36)$$

c'est-à-dire que la puissance de la stratégie basée sur le critère de Neumann-Pearson [cf. (1.26)], égale à

$$\int_{G_1}\cdots\int W_n(x_1, \ldots, x_n \mid s_1)\, dx_1 \ldots dx_n = 1-\beta = 1 - F_{11}(c), \qquad (1.37)$$

est supérieure à la puissance pour toute autre stratégie satisfaisant à la condition (1.34).

En comparant (1.32) et (1.20) on voit que la stratégie basée sur le critère de Neumann-Pearson est un *cas particulier* de la décision de Bayes où la grandeur μc^* est remplacée par c trouvé de (1.33).

Ainsi, tous les critères de qualité envisagés ci-dessus conduisent à des procédures semblables de prise de décision : d'après l'échantillon $x_1, \ldots, x_n$ on calcule le rapport de vraisemblance $l(x_1, \ldots, x_n)$ et l'on accepte ou l'on rejette l'hypothèse H_0 suivant que cette grandeur se trouve au-dessous ou au-dessus d'un certain seuil, établi à l'avance conformément au critère adopté.

1.2.5. Méthode de calcul des probabilités conditionnelles des erreurs. Si $\Phi(y)$ est une fonction monotone, comparer $l(x_1, \ldots, x_n)$ au seuil c *équivaut* à comparer $\Phi[l(x_1, \ldots, x_n)]$ au seuil $\Phi(c)$. La procédure de vérification de l'hypothèse reste optimale. Les probabilités conditionnelles des erreurs sont calculées à l'aide des formules du type (1.25), (1.26) où les distributions des rapports de vraisemblance sont remplacées par les distributions de $\Phi(l)$ et la borne d'intégration c est remplacée par la grandeur $\Phi(c)$.

Lorsque l'échantillon se compose d'éléments indépendants, la fonction de vraisemblance de l'échantillon est le produit de leurs densités de probabilité unidimensionnelles [cf. (1.6)]. Une telle forme de la fonction de vraisemblance implique le choix de la fonction Φ sous la forme du *logarithme* du rapport de vraisemblance. Il se trouve possible d'écrire le logarithme du rapport multidimensionnel de vraisemblance comme une somme de logarithmes de rapports unidimensionnels de vraisemblance :

$$\ln l(x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln l(x_i) = \sum_{i=1}^{n} \ln \frac{w_1(x_i \mid s_1)}{w_1(x_i \mid s_0)}. \qquad (1.38)$$

Si l'échantillon $(x_1, \ldots, x_n)$ est homogène, $\ln l(x_1, \ldots, x_n)$ est une somme de grandeurs aléatoires *indépendantes* $\ln l(x_i)$ ayant même distribution. Désignons par $\Theta_1(v \mid s_0)$, $\Theta_1(v \mid s_1)$ les fonctions caractéristiques conditionnelles de la grandeur aléatoire $\ln l(x)$ respectivement pour les hypothèses H_0 et H_1. En vertu de la relation bien connue en théorie des probabilités [cf. (3.115), tome I)

$$\Theta_1(v \mid s_0) = \int_{-\infty}^{\infty} w_1(x \mid s_0) \exp\left[iv \ln \frac{w_1(x \mid s_1)}{w_1(x \mid s_0)}\right] dx, \qquad (1.39)$$

$$\Theta_1(v \mid s_1) = \int_{-\infty}^{\infty} w_1(x \mid s_1) \exp\left[iv \ln \frac{w_1(x \mid s_1)}{w_1(x \mid s_0)}\right] dx, \qquad (1.40)$$

les densités de probabilité des variables aléatoires $\ln l(x_1, \ldots, x_n)$ sont

$$W_1(z \mid s_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Theta_1(v \mid s_0)]^n e^{-ivz} \, dv, \qquad (1.41)$$

$$W_1(z \mid s_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Theta_1(v \mid s_1)]^n e^{-ivz} \, dv. \qquad (1.42)$$

On peut maintenant, en utilisant (1.25) et (1.26), trouver les probabilités conditionnelles de décisions erronées à l'aide des formules suivantes

$$\alpha = \int_{\ln c}^{\infty} W_1(z \mid s_0)\, dz = 1 - F_1(\ln c \mid s_0), \tag{1.43}$$

$$\beta = \int_{-\infty}^{\ln c} W_1(z \mid s_1)\, dz = F_1(\ln c \mid s_1), \tag{1.44}$$

où $F_1(z \mid s_0)$ et $F_1(z \mid s_1)$ sont les fonctions de répartition conditionnelles de $\ln l(x_1, \ldots, x_n)$.

Si n est grand et les variances conditionnelles $M_2\{\ln l(x) \mid s_i\}$ sont bornées, la loi de distribution de $\ln l(x_1, \ldots, x_n)$ en tant que somme d'un grand nombre de variables aléatoires indépendantes de même distribution [cf. (1.38)] tend, en vertu du théorème de Liapounov (cf. § 3.4, tome I), vers une loi normale. Dans ce cas

$$W_1(z \mid s_0) \sim \frac{1}{\sqrt{2\pi n M_{20}}} \exp\left\{-\frac{(z - n m_{10})^2}{2 n M_{20}}\right\}, \tag{1.45}$$

$$W_1(z \mid s_1) \sim \frac{1}{\sqrt{2\pi n M_{21}}} \exp\left\{-\frac{(z - n m_{11})^2}{2 n M_{21}}\right\}, \tag{1.46}$$

où

$$m_{10} = m_1\{\ln l(x) \mid s_0\},\quad M_{20} = M_2\{\ln l(x) \mid s_0\}, \tag{1.47}$$
$$m_{11} = m_1\{\ln l(x) \mid s_1\},\quad M_{21} = M_2\{\ln l(x) \mid s_1\}. \tag{1.48}$$

On a alors les relations *asymptotiques* suivantes [cf. (1.43) et (1.44)]:

$$\alpha \sim 1 - F\left(\frac{\ln c - n m_{10}}{\sqrt{n M_{20}}}\right), \tag{1.49}$$

$$\beta \sim F\left(\frac{\ln c - n m_{11}}{\sqrt{n M_{21}}}\right), \tag{1.50}$$

où $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$ est l'intégrale de Laplace (cf. § 2.2.2, tome I).

On peut montrer (voir problème 1.10) que $m_{10} < 0$ et $m_{11} > 0$. Il s'ensuit que pour $n \to \infty$ on a simultanément $\alpha \to 0$ et $\beta \to 0$ (pour le critère de Neumann-Pearson, α étant donné, β tend vers zéro).

Une stratégie pour laquelle la probabilité d'erreur tend asymptotiquement vers zéro (pour $n \to \infty$) est dite *consistante*.

1.2.6. Stratégie du minimax. Nous allons également étudier la stratégie du minimax en utilisant le fait que celle-ci est un cas particulier de la stratégie de Bayes pour la distribution a priori des états la moins favorable (cf. § 1.1.6). Comme dans le cas qui nous intéresse de vérification des hypothèses simples il y a seulement deux états possibles s_0 et s_1, la distribution mentionnée est donnée par la

seule probabilité q (ou $p = 1 - q$). Ainsi, pour trouver celle des valeurs de q à laquelle correspond la plus grande valeur du risque de Bayes, il faut trouver le maximum de la grandeur R^* en tant que fonction de q en utilisant (1.23), sans oublier que cette variable entre dans le second membre de (1.23) tant directement que par l'intermédiaire de α^* et β^* [cf. (1.25) et (1.26)]. En dérivant le second membre de (1.23) par rapport à q, en annulant cette dérivée et en tenant compte du fait que les points de la surface séparant les domaines critique et admissible de l'espace des échantillons satisfont à la condition (1.21), on obtient une équation transcendante par rapport à la valeur cherchée la moins favorable de la probabilité q:

$$\Pi_{00}[1 - \alpha^*(q)] + \Pi_{01}\alpha^*(q) = \Pi_{10}\beta^*(q) + \Pi_{11}[1 - \beta^*(q)]. \tag{1.51}$$

L'équation (1.51) découle directement de (1.17) et (1.17') car elle exprime simplement l'égalité des risques conditionnels $r_0 = r_1$ pour la stratégie de Bayes. Cette égalité, comme nous l'avons mentionné au § 1.1.6, caractérise justement la stratégie du minimax. En portant (1.25) et (1.26) dans (1.51) on peut écrire sous la forme suivante l'équation transcendante par rapport à q $\left(\text{ou à } \mu = \frac{q}{p}\right)$

$$\frac{\Pi_{01} - \Pi_{00}}{\Pi_{10} - \Pi_{11}} + c^*[1 - F_{10}(\mu c^*)] = F_{11}(\mu c^*). \tag{1.51'}$$

En résolvant l'équation (1.51') et en prenant sa racine μ_{mm} (et par conséquent q_{mm}) à laquelle correspond le maximum absolu du risque de Bayes, on arrive à la *stratégie du minimax* suivante: on adopte la décision γ_1 (rejet de l'hypothèse H_0) si pour l'échantillon observé $x_1, \ldots, x_n$ on a l'inégalité

$$\frac{p_{mm}}{q_{mm}} l(x_1, \ldots, x_n) \geqslant c^*, \tag{1.52}$$

et la décision γ_0 (affirmant l'authenticité de l'hypothèse H_0) si c'est l'inégalité contraire de (1.52) qui se trouve vérifiée.

En vertu de (1.23), (1.25) et (1.26) la grandeur du risque du minimax peut être calculée à l'aide de la formule

$$R_{mm} = q_{mm}\Pi_{00} + p_{mm}\Pi_{10} + q_{mm}(\Pi_{01} - \Pi_{00})\alpha_{mm} - \\ - p_{mm}(\Pi_{10} - \Pi_{11})(1 - \beta_{mm}), \tag{1.53}$$

où

$$\alpha_{mm} = 1 - F_{10}(\mu_{mm}c^*), \quad \beta_{mm} = F_{11}(\mu_{mm}c^*). \tag{1.54}$$

La différence $R_{mm} - R^* \geqslant 0$ entre le risque du minimax (pour q inconnu) et le risque de Bayes (pour q donné) se trouve être le prix qu'il faut payer pour l'absence d'information a priori concernant les états de l'effet étudié.

Résumant ce qui a été dit aux §§ 1.2.2 à 1.2.5 on peut noter que pour tous les critères envisagés la procédure de vérification d'une

hypothèse simple revient à comparer le rapport de vraisemblance au seuil c *). Les expressions donnant ce seuil, pour les différents critères, sont données dans la table 1.

Table 1

Critère	Seuil c
de Bayes	$\frac{q}{p}\frac{\Pi_{01}-\Pi_{00}}{\Pi_{10}-\Pi_{11}}$
du maximum de probabilité a posteriori	$\mu=\frac{q}{p}$
du maximum de vraisemblance	1
de Neumann-Pearson	solution de l'équation $F_{10}(c)=1-\alpha$
du minimax	$\mu_{mm}\frac{\Pi_{01}-\Pi_{00}}{\Pi_{10}-\Pi_{11}}$, μ_{mm} de (1.51)

1.2.7. Echantillonnage progressif. La particularité essentielle de toutes les méthodes exposées ci-dessus de choix d'une décision est que l'échantillon était de taille constante n. Il y a une autre approche du problème de choix d'une stratégie où l'on renonce à prendre un échantillon de taille donnée, celle-ci étant définie au cours de l'expérience même, suivant les résultats des observations déjà effectuées. On commence par observer une seule valeur x_1 (extraction d'un échantillon de taille $n = 1$) et à partir de cette valeur, suivant une stratégie établie à l'avance on adopte l'une des trois décisions suivantes:

1) on accepte l'hypothèse H_0 (décision γ_0) ; 2) on rejette l'hypothèse H_0, c'est-à-dire que l'on accepte l'hypothèse H_1 (décision γ_1) ; 3) on *poursuit les observations*, c'est-à-dire qu'on renonce à adopter l'une des décisions γ_0 ou γ_1.

Si l'on adopte la décision γ_0 ou γ_1 l'expérience se trouve terminée. Dans le cas contraire on prend l'échantillon suivant et l'on répète la même procédure: en observant l'échantillon (x_1, x_2) de taille $n = 2$, on accepte l'hypothèse H_0, ou on la rejette, ou bien on considère que cet échantillon est insuffisant pour pouvoir définitivement adopter l'une des décisions γ_0 ou γ_1. Si l'on accepte ou l'on rejette l'hypothèse H_0, l'expérience se trouve par là même terminée, et si la

*) Notons que le domaine critique pour lequel $l(x_1, \ldots, x_n) \geqslant c$ n'est pas forcément *simplement connexe* (cf. problème 1.5).

décision définitive n'est pas prise on prélève l'élément suivant x_3, puis l'on répète la même procédure pour l'échantillon (x_1, x_2, x_3).

L'essai se termine sur l'échantillon qui permet de prendre l'une des décisions γ_0 ou γ_1. L'échantillonnage progressif a ceci de particulier que la taille de l'échantillon n'est pas fixée à l'avance, c'est une *variable aléatoire.* L'espace des échantillons du nombre correspondant de dimensions doit maintenant être divisé non pas en deux mais en trois domaines: le domaine critique G_1, le domaine admissible G_0 et le domaine intermédiaire G_{int}. C'est la partition de l'espace des échantillons en trois domaines qui indique si l'on doit adopter l'une des décisions γ_0 ou γ_1, ou si l'on doit poursuivre l'expérience. Si la valeur échantillonnée tombe dans le domaine critique G_1 l'hypothèse H_0 est rejetée; si elle tombe dans le domaine admissible G_0 l'hypothèse H_0 est adoptée, et si elle tombe dans le domaine intermédiaire G_{int}, il y a lieu de poursuivre l'expérience.

Tout comme dans le cas des méthodes d'échantillonnage non progressif, l'espace des échantillons peut être divisé d'une infinité de manières. Par conséquent, des stratégies les plus diverses sont possibles et évidemment, il faut de nouveau des critères de qualité permettant de comparer les différentes procédures d'échantillonnage progressif et d'en choisir la meilleure. Il est raisonnable et naturel d'adopter comme critère un *prix moyen minimal de l'expérience.* Si l'on considère que le prix de l'expérience est proportionnel à la taille de l'échantillon n, le critère de qualité de la stratégie d'échantillonnage progressif peut être le *minimum de la taille moyenne* de l'échantillon indispensable pour que l'une des décisions γ_0 ou γ_1 soit prise, à condition que le niveau de signification ne dépasse pas α et que la puissance ne soit pas inférieure à $1 - \beta$. Il faut noter que dans ce cas les tailles moyennes de l'échantillon $m_1 \{n \mid H_0\}$ et $m_1 \{n \mid H_1\}$ pour les hypothèses H_0 et H_1 respectivement ne sont pas en général égales entre elles et doivent être *toutes deux* minimisées.

A. Wald [3] a montré que parmi toutes les stratégies (d'échantillonnage progressif ou non) pour lesquelles les probabilités conditionnelles des erreurs ne dépassent pas les grandeurs α et β, c'est la stratégie d'échantillonnage progressif dans laquelle on compare le *rapport de vraisemblance* $l(x_1, \ldots, x_n)$ aux deux seuils *) c_0 et c_1 qui donne les plus petites valeurs de $m_1 \{n \mid H_0\}$ et $m_1 \{n \mid H_1\}$.

La partition optimale de l'espace des échantillons est donnée par les inégalités suivantes:

pour le domaine admissible G_0

$$c_0 < l(x_1, \ldots, x_k) < c_1, \quad k = 1, \ldots, n-1, \tag{1.55}$$
$$l(x_1, \ldots, x_n) \leqslant c_0;$$

*) Les valeurs de c_0 et c_1 dépendent en général de n.

pour le domaine critique G_1

$$c_0 < l(x_1, \ldots, x_k) < c_1, \; k = 1, \ldots, n-1, \tag{1.55'}$$
$$l(x_1, \ldots, x_n) \geqslant c_1;$$

pour le domaine intermédiaire G_{int}

$$c_0 < l(x_1, \ldots, x_k) < c_1, \; k = 1, \ldots, n. \tag{1.55''}$$

On rencontre de grosses difficultés mathématiques lors du calcul exact des seuils c_0 et c_1. Cependant, il a été démontré dans [3] que

$$c_1 \leqslant \max\left(\frac{\beta}{1-\alpha}, \; \frac{1-\beta}{\alpha}\right), \tag{1.56}$$

$$c_0 \geqslant \min\left(\frac{\beta}{1-\alpha}, \; \frac{1-\beta}{\alpha}\right). \tag{1.56'}$$

Pratiquement dans les cas les plus fréquents et donc les plus intéressants, quand les probabilités conditionnelles d'erreurs ne dépassent pas la valeur 0,5, on a $\frac{1-\beta}{\alpha} > \frac{\beta}{1-\alpha}$, et ces inégalités peuvent être écrites sous la forme suivante:

$$c_1 \leqslant \frac{1-\beta}{\alpha}, \quad c_0 \geqslant \frac{\beta}{1-\alpha}. \tag{1.57}$$

Puis on remplace les inégalités (1.57) par des égalités, ce qui n'augmente pas sensiblement l'erreur (pour plus de détail voir [7], ch. 3).

Ainsi, dans une certaine approximation, la stratégie d'échantillonnage progressif peut être formulée de la manière suivante: lors de la n-ième observation on adopte la décision γ_0 si *)

$$\frac{\beta}{1-\alpha} < l(x_1, \ldots, x_k) < \frac{1-\beta}{\alpha}, \quad k = 1, \ldots, n-1 \tag{1.58}$$

et

$$l(x_1, \ldots, x_n) \leqslant \frac{\beta}{1-\alpha} < 1; \tag{1.58'}$$

on adopte la décision γ_1 si les inégalités (1.58) se trouvent vérifiées et si

$$l(x_1, \ldots, x_n) \geqslant \frac{1-\beta}{\alpha} > 1. \tag{1.58''}$$

Notons que pour les approximations faites ci-dessus, à la différence de la stratégie de Bayes, dans celle d'échantillonnage progressif on compare le rapport de vraisemblance à des seuils qui ne dépendent pas des probabilités a priori des états ni des valeurs des pertes. Ces seuils sont donnés par les probabilités des erreurs de première et de seconde espèce.

*) La probabilité que le nombre n d'étapes de la procédure d'échantillonnage progressif de vérification des hypothèses *ne soit pas limité* est nulle (voir [7]).

Si au lieu du rapport de vraisemblance on calcule son logarithme, la stratégie mentionnée peut s'écrire de la manière suivante: pour n observations on adopte la décision γ_0 si

$$\ln \frac{\beta}{1-\alpha} < \ln l(x_1, \ldots, x_k) < \ln \frac{1-\beta}{\alpha}, \quad k=1, \ldots, n-1 \qquad (1.59)$$

et

$$\ln l(x_1, \ldots, x_n) \leqslant \ln \frac{\beta}{1-\alpha} < 0; \qquad (1.59')$$

on adopte la décision γ_1 si en plus de (1.59) on a

$$\ln l(x_1, \ldots, x_n) \geqslant \ln \frac{1-\beta}{\alpha} > 0. \qquad (1.59'')$$

Calculons maintenant les valeurs moyennes de la taille de l'échantillon $m_1\{n \mid H_0\}$ et $m_1\{n \mid H_1\}$ correspondant à la stratégie d'échantillonnage progressif. Dans le cas envisagé le logarithme du rapport de vraisemblance est égal à la somme d'un *nombre aléatoire* de variables aléatoires [cf. (1.38)]

$$\ln l(x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln l(x_i).$$

Ainsi [cf. (3.140), tome I]

$$m_1 \{\ln l(x_1, \ldots, x_n) \mid H_{0;\,1}\} = m_1 \{n \mid H_{0;\,1}\}\, m_1 \{\ln l(x) \mid H_{0;\,1}\},$$

d'où

$$m_1\{n \mid H_0\} = \frac{m_1\{\ln l(x_1, \ldots, x_n) \mid H_0\}}{m_1\{\ln l(x) \mid H_0\}}, \qquad (1.60)$$

$$m_1\{n \mid H_1\} = \frac{m_1\{\ln l(x_1, \ldots, x_n) \mid H_1\}}{m_1\{\ln l(x) \mid H_1\}}. \qquad (1.60')$$

Supposons que lorsque l'on prend une décision (γ_0 ou γ_1) au n-ième pas, la valeur du rapport de vraisemblance *coïncide exactement* avec l'un des seuils c_0 ou c_1 (c'est-à-dire que l'on négligera l'intersection du seuil lors de l'étape finale de la vérification de l'hypothèse). Dans ce cas $\ln l(x_1, \ldots, x_n)$ est une variable aléatoire discrète pouvant prendre deux valeurs: $\ln c_0$ et $\ln c_1$ avec les probabilités $1-\alpha$ et α si c'est l'hypothèse H_0 qui est vraie et avec les probabilités β et $1-\beta$ si c'est l'hypothèse H_1 qui est vraie. Par conséquent on a

$$m_1\{\ln l(x_1, \ldots, x_n) \mid H_0\} = (1-\alpha) \ln \frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha}, \qquad (1.61)$$

$$m_1\{\ln l(x_1, \ldots, x_n) \mid H_1\} = \beta \ln \frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha}. \qquad (1.61')$$

En portant (1.61) dans (1.60) et (1.61′) dans (1.60′) on obtient

$$m_1\{n \mid H_0\} = \frac{(1-\alpha)\ln\frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha}}{m_{10}}, \qquad (1.62)$$

$$m_1\{n \mid H_1\} = \frac{\beta \ln\frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha}}{m_{11}}, \qquad (1.62')$$

où [cf. (1.47) et (1.48)]

$$m_{10} = m_1\{\ln l(x) \mid H_0\} = \int_{-\infty}^{\infty} W_1(x \mid s_0) \ln \frac{W_1(x \mid s_1)}{W_1(x \mid s_0)}\, dx\,; \qquad (1.63)$$

$$m_{11} = m_1\{\ln l(x) \mid H_1\} = \int_{-\infty}^{\infty} W_1(x \mid s_1) \ln \frac{W_1(x \mid s_1)}{W_1(x \mid s_0)}\, dx, \qquad (1.63')$$

avec $m_{10} < 0$, $m_{11} > 0$, ce qui découle directement de la formule (23) du problème 1.10.

Les valeurs moyennes de la taille de l'échantillon données par (1.62) et (1.62′) sont *minimales* parmi toutes les valeurs *possibles* pour toutes les autres stratégies (y compris celles d'*échantillonnage non progressif*) et garantissent des probabilités d'erreurs limitées à des valeurs données. C'est là l'avantage essentiel de l'échantillonnage progressif car en moyenne il permet d'économiser tant sur la durée que sur le coût de l'expérience. Cependant il faut noter qu'il s'agit *d'une économie en moyenne* ce qui est très important. En effet, dans le cas d'échantillonnage progressif, la taille n de l'échantillon est une variable aléatoire qui peut prendre des valeurs nettement supérieures à la valeur moyenne et il peut se faire que dans certains cas la procédure optimale d'échantillonnage progressif conduit à un échantillon trop grand.

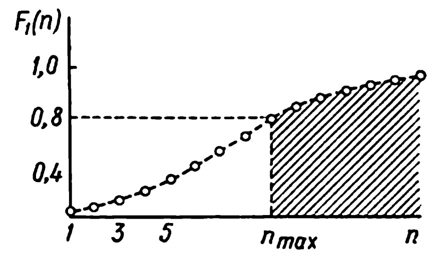


Fig. 1.3. Fonction de répartition de la taille d'un échantillon lors de l'échantillonnage progressif

C'est ce que l'on peut voir sur la figure 1.3. La partie hachurée correspond à des échantillons de taille supérieure à une certaine valeur acceptable n_{max}. Bien que dans 80 % des cas la procédure de vérification des hypothèses soit de moindre durée par rapport à celle des méthodes de taille fixe $n = n_{max}$, il y aura quand même des cas « malheureux » où l'analyse par la méthode d'échantillonnage progressif se trouve plus longue. Il est naturel d'essayer de parer à l'in-

convénient, pour cela on se donne à l'avance la valeur maximale de la taille $n_{\max}$ de l'échantillon. Une fois cette valeur atteinte, on arrête l'expérience et l'on prend obligatoirement l'une des décisions γ_0 ou γ_1 (si, évidemment, on n'a pas pris une décision avant). On évite ainsi les cas où $n > n_{\max}$.

La procédure mentionnée de vérification des hypothèses est appelée plan d'échantillonnage *tronqué*. Ici, avant d'atteindre $n_{\max}$, on compare le rapport de vraisemblance correspondant à deux seuils, et lorsque la taille de l'échantillon atteint la valeur $n_{\max}$, le rapport de vraisemblance correspondant est comparé non pas à deux seuils mais à un seul, comme lors de l'analyse d'après les plans de taille fixe. Plus $n_{\max}$ est petit, c'est-à-dire plus le plan est tronqué, moindre sera le gain en temps moyen obtenu avec l'échantillonnage progressif.

La stratégie dans le cas d'échantillonnage progressif tronqué se formule comme suit: si pour la taille de l'échantillon $n \leqslant n_{\max}$ la stratégie (1.59) ne conduit pas à une décision (γ_0 ou γ_1), l'hypothèse H_0 est rejetée (on adopte la décision γ_1) si

$$\ln \frac{\beta}{1-\alpha} < \ln c \leqslant \ln l(x_1, \ldots, x_{n_{\max}}) < \ln \frac{1-\beta}{\alpha}, \qquad (1.64)$$

et on accepte l'hypothèse H_0 (décision γ_0) si

$$\ln \frac{\beta}{1-\alpha} < \ln l(x_1, \ldots, x_{n_{\max}}) < \ln c < \ln \frac{1-\beta}{\alpha}. \qquad (1.64')$$

En utilisant la règle (1.64) on doit s'attendre à ce que les probabilités d'erreurs α_{tr}, β_{tr} de première et de seconde espèce s'avèrent parfois supérieures aux erreurs données: $\alpha_{tr} \geqslant \alpha$; $\beta_{tr} \geqslant \beta$, car les décisions erronées adoptées pouvaient ne pas apparaître si l'on continuait les essais ($n > n_{\max}$). Les limites supérieures des probabilités conditionnelles des erreurs pour le plan d'échantillonnage progressif tronqué sont données par les inégalités

$$\alpha_{tr} \leqslant \alpha + P\left\{\ln \frac{1-\beta}{\alpha} > \ln l(x_1, \ldots, x_{n_{\max}}) \geqslant \ln c \mid H_0\right\}, \qquad (1.65)$$

$$\beta_{tr} \leqslant \beta + P\left\{\ln c > \ln l(x_1, \ldots, x_{n_{\max}}) > \ln \frac{\beta}{1-\alpha} \middle| H_1\right\}. \qquad (1.65')$$

Pour $n_{\max} \gg 1$, comme il a été démontré dans [3], ces inégalités deviennent

$$\alpha_{tr} \leqslant \alpha + F\left[\frac{\ln \frac{1-\beta}{\alpha} - n_{\max} m_{10}}{\sqrt{n_{\max} M_{20}}}\right] - F\left[\frac{\ln c - n_{\max} m_{10}}{\sqrt{n_{\max} M_{20}}}\right],$$

$$\beta_{tr} \leqslant \beta + F\left[\frac{\ln c - n_{\max} m_{11}}{\sqrt{n_{\max} M_{21}}}\right] - F\left[\frac{\ln \frac{\beta}{1-\alpha} - n_{\max} m_{11}}{\sqrt{n_{\max} M_{21}}}\right],$$

où m_{10}, m_{11}, M_{20}, M_{21} sont donnés par (1.47) et (1.48) et $F(x)$ est l'intégrale de Laplace.

Notons que comme $\ln \frac{\beta}{1-\alpha} < 0$ et $\ln \frac{1-\beta}{\alpha} > 0$, la valeur du seuil $c = 1$ pour le plan d'échantillonnage progressif est toujours acceptable [cf. (1.64), (1.64')]. A l'heure actuelle on ne peut pas donner de recommandations plus ou moins fondées quant à la valeur optimale du seuil c.

1.3. VÉRIFICATION DE L'HYPOTHÈSE SIMPLE CONCERNANT LE PARAMÈTRE D'UNE DISTRIBUTION

1.3.1. Valeur moyenne d'une variable normalement distribuée. Nous allons maintenant donner des exemples simples permettant d'illustrer les méthodes générales exposées plus haut. Commençons par l'exemple de vérification de l'hypothèse H_0 en vertu de laquelle la valeur moyenne d'une variable normalement distribuée est égale à a_0, avec l'alternative simple H_1 que cette valeur moyenne est a_1. Supposons que l'on connaît exactement la variance σ^2 d'une variable aléatoire normale et que l'échantillon sur lequel on vérifie l'hypothèse se compose d'éléments indépendants $x_1, \ldots, x_n$. Comme toutes les procédures optimales de vérification des hypothèses que nous avons envisagées se réduisent à comparer le rapport de vraisemblance $l(x_1, \ldots, x_n)$, ou son logarithme, à un certain seuil (ou à deux seuils dans le plan d'échantillonnage progressif), nous écrivons tout d'abord l'expression de $\ln l(x_1, \ldots, x_n)$ pour la loi normale lorsque les éléments de l'échantillon sont indépendants:

$$\ln l(x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln l(x_i) =$$

$$= \sum_{i=1}^{n} \ln \frac{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-a_1)^2}{2\sigma^2}\right]}{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-a_0)^2}{2\sigma^2}\right]} = \frac{1}{2\sigma^2} \sum_{i=1}^{n} [(x_i-a_0)^2 -$$

$$-(x_i-a_1)^2] = \frac{a_1-a_0}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{n(a_1^2-a_0^2)}{2\sigma^2}. \quad (1.66)$$

On peut maintenant facilement formuler la stratégie pour un plan de taille n fixe de l'échantillon : on adopte la décision γ_1 (la moyenne est égale à a_1) si pour l'échantillon observé on a

$$\frac{1}{n} \sum_{i=1}^{n} x_i \geqslant \frac{a_0+a_1}{2} + \frac{\sigma^2 \ln c}{n(a_1-a_0)} = K, \quad a_1 > a_0, \quad (1.67)$$

et l'on adopte la décision γ_0 (la moyenne est égale à a_0) si on a l'inégalité contraire de (1.67). La grandeur du seuil pour les différents critères est donnée dans la table 1, page 35.

Si $a_1 < a_0$, l'inégalité (1.67) change de sens et la condition choix de la décision γ_1 s'écrit sous la forme

$$\frac{1}{n}\sum_{i=1}^{n} x_i \leqslant \frac{a_0+a_1}{2} - \frac{\sigma^2 \ln c}{n\,(a_0-a_1)}, \quad a_1 < a_0. \tag{1.67'}$$

Ainsi, la procédure de vérification de l'hypothèse concernant la valeur moyenne d'une variable aléatoire normale se réduit à comparer la moyenne des variables échantillonnées au seuil

$$K = \frac{a_0+a_1}{2} + \frac{\sigma^2 \ln c}{n\,(a_1-a_0)}. \tag{1.68}$$

La surface séparant le domaine admissible et le domaine critique de l'espace des échantillons est dans ce cas un hyperplan perpendiculaire au vecteur unité et éloigné de nK de l'origine des coordonnées.

Il est alors facile de trouver les expressions des probabilités conditionnelles d'erreurs. Comme la somme de variables aléatoires normales est elle aussi une variable normale, et, étant donné que

$$m_1\left\{\frac{1}{n}\sum_{i=1}^{n} x_i \mid H_0\right\} = a_0, \quad m_1\left\{\frac{1}{n}\sum_{i=1}^{n} x_i \mid H_1\right\} = a_1, \tag{1.69}$$

et si les éléments de l'échantillon sont indépendants

$$M_2\left\{\frac{1}{n}\sum_{i=1}^{n} x_i \mid H_0\right\} = M_2\left\{\frac{1}{n}\sum_{i=1}^{n} x_i \mid H_1\right\} = \frac{\sigma^2}{n}, \tag{1.70}$$

en utilisant (1.67) [cf. également (1.43) et (1.44)], on obtient pour $a_1 > a_0$

$$\alpha = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \int_{K}^{\infty} e^{-\frac{n(y-a_0)^2}{2\sigma^2}}\, dy =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{K-a_0}{\sigma}\sqrt{n}}^{\infty} e^{-\frac{t^2}{2}}\, dt = 1 - F\left(\frac{K-a_0}{\sigma}\sqrt{n}\right), \tag{1.71}$$

$$\beta = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{K} e^{-\frac{n(y-a_1)^2}{2\sigma^2}}\, dy =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{K-a_1}{\sigma}\sqrt{n}} e^{-\frac{t^2}{2}}\, dt = F\left(\frac{K-a_1}{\sigma}\sqrt{n}\right), \tag{1.72}$$

où $F(x)$ est l'intégrale de Laplace [cf. (1.49) et (1.50)].

La figure 1.4 illustre cette stratégie. Les courbes en pointillé correspondent aux densités de probabilité associées aux hypothèses H_0 et H_1. Les courbes en trait continu donnent les densités de probabilité de la moyenne arithmétique des valeurs échantillonnées pour les mêmes hypothèses. Le seuil est établi à partir du critère du

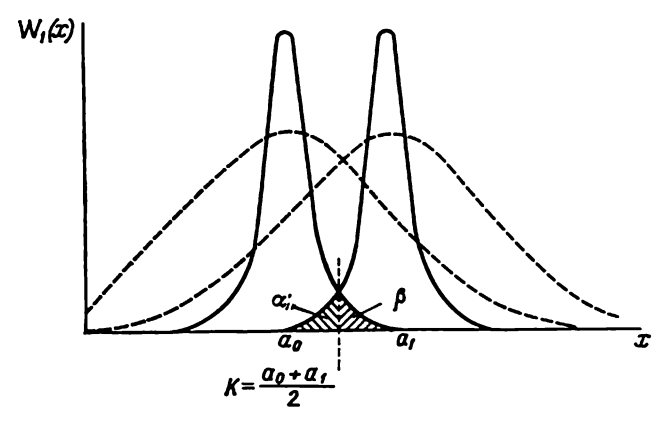


Fig. 1.4. Stratégie de vérification de l'hypothèse sur la moyenne d'une variable aléatoire normale

maximum de vraisemblance ($c = 1$). Les aires des parties hachurées sont égales aux probabilités des erreurs de première et de seconde espèce qui sont identiques pour $K = \frac{a_0+a_1}{2}$.

Dans les formules (1.71) et (1.72) la grandeur K pour les trois premiers critères figurant à la table 1 est donnée directement par la formule (1.68), ainsi ces formules deviennent

$$\alpha = 1 - F\left[\frac{a_1-a_0}{2\sigma}\sqrt{n} + \frac{\sigma \ln c}{\sqrt{n}\,(a_1-a_0)}\right], \quad (1.73)$$

$$\beta = F\left[-\frac{a_1-a_0}{2\sigma}\sqrt{n} + \frac{\sigma \ln c}{\sqrt{n}\,(a_1-a_0)}\right], \quad a_1 > a_0. \quad (1.73')$$

Les formules ci-dessus montrent que les probabilités conditionnelles d'erreurs pour c donné dépendent encore de la *seule* grandeur $d_n = \frac{a_1-a_0}{\sigma}\sqrt{n}$. Il est facile de voir que la moyenne et la variance du logarithme du rapport de vraisemblance sont liées à cette grandeur par des relations simples

$$m_1\{\ln l(x_1, \ldots, x_n) \mid H_1\} = -m_1\{\ln l(x_1, \ldots, x_n) \mid H_0\} = \frac{d_n^2}{2},$$

$$M_2\{\ln l(x_1, \ldots, x_n) \mid H_1\} = M_2\{\ln l(x_1, \ldots, x_n) \mid H_0\} = d_n^2.$$

Notons qu'en vertu de (1.49) et (1.50) pour $n \to \infty$ les probabilités d'erreurs tendent vers zéro conformément au résultat obtenu au

§ 1.2.5. Dans ce cas la distribution de la moyenne arithmétique des valeurs échantillonnées tend vers une fonction delta aux points $y = a_0$, $y = a_1$, et le seuil $K \to \frac{a_0 + a_1}{2}$.

Pour le critère du *maximum de vraisemblance* (fig. 1.4)

$$\alpha = \beta = 1 - F\left(\frac{a_1 - a_0}{2\sigma}\sqrt{n}\right). \tag{1.74}$$

Pour $\alpha = \beta$ donné, on obtient à partir de (1.74) la taille requise de l'échantillon

$$n = 4\left(\frac{\sigma}{a_1 - a_0}\right)^2 x_\alpha^2, \tag{1.74'}$$

où

$$x_\alpha = \arg F(x).$$

En mathématiques statistiques la grandeur x_α est généralement appelée quantile d'une variable aléatoire *), caractérisé par l'abscisse de la courbe de distribution telle que la partie de l'aire sous cette courbe se trouvant à droite du point x_α soit égale à α, autrement dit

$$P\{\xi \geqslant x_\alpha\} = \alpha. \tag{1.75}$$

Dans la formule (1.74') la grandeur x_α est ainsi le quantile d'une variable aléatoire normalement distribuée. On a donné dans l'annexe VII une table des quantiles de la loi normale.

Pour le critère de Neumann-Pearson, ayant fixé le niveau de signification α, la grandeur K est donnée par la formule ($a_1 > a_0$)

$$K = a_0 + \frac{\sigma}{\sqrt{n}} x_\alpha. \tag{1.76}$$

Dans ce cas la probabilité d'erreur de seconde espèce peut être obtenue en portant (1.76) dans (1.72):

$$\beta = F\left[-\frac{a_1 - a_0}{\sigma}\sqrt{n} + x_\alpha\right] \tag{1.77}$$

ou

$$x_\alpha - x_{1-\beta} = \frac{a_1 - a_0}{\sigma}\sqrt{n}, \quad a_1 > a_0. \tag{1.77'}$$

La formule (1.77') établit la correspondance entre la taille n de l'échantillon, la probabilité conditionnelle d'erreur α de première

*) Il est évident que le quantile de la variable aléatoire ξ coïncide avec le quantile $1 - \alpha$ de sa fonction de répartition $F_1(x)$, car en vertu de (1.75) x_α est la racine de l'équation $F_1(x) = 1 - \alpha$ (cf. tome I, § 2.1.3). Des cas particuliers peuvent apparaître lorsque $F(x)$ est discontinue ou lorsque sur l'axe réel il existe des intervalles où la fonction de répartition reste constante. Pour éviter toute confusion il faut noter que parfois (cf., par exemple, [5]) le quantile peut être défini autrement: dans (1.75) on remplace ξ par sa valeur absolue de sorte qu'on obtient $P\{|\xi| \geqslant \lambda_\alpha\} = \alpha$. Dans cet ouvrage on utilise partout la définition (1.75). Pour une variable aléatoire dont la fonction de répartition est symétrique par rapport à l'origine des coordonnées on a $x_\alpha = \lambda_{2\alpha}$.

espèce et la probabilité minimale possible d'erreur β de seconde espèce. Pour $n \to \infty$ en vertu de (1.77) on a $\beta \to 0$, ceci conformément au résultat général du § 1.2.5.

Remarquons que le seuil K déterminé par le critère de Neumann-Pearson d'après (1.76) *ne dépend pas* de a_1.

Il faut souligner que les probabilités d'erreur α et β dans la stratégie de Bayes et la probabilité d'erreur de seconde espèce (pour α donné) pour le critère de Neumann-Pearson ne dépendent pas de chacune des grandeurs n, a_0, a_1 et σ séparément, mais seulement du rapport $d_n = \frac{a_1-a_0}{\sigma}\sqrt{n}$. Par conséquent, lorsque la grandeur $\frac{a_1-a_0}{\sigma}$ diminue de K fois (distinction entre deux hypothèses voisines), pour que les probabilités d'erreur soient les mêmes il faut que la taille de l'échantillon augmente de K^2 fois. La stratégie ne sera consistante que si $\frac{a_1-a_0}{\sigma}$ décroît plus lentement que $\frac{1}{\sqrt{n}}$ pour $n \to \infty$.

Si $a_1 < a_0$, la décision γ_1 suivant le critère de Neumann-Pearson sera adoptée à condition que [comparer avec (1.67')]

$$\frac{1}{n}\sum_{i=1}^{n} x_i \leqslant a_0 - \frac{\sigma}{\sqrt{n}}\, x_\alpha .$$

La probabilité d'erreur de seconde espèce sera alors égale à

$$\beta = F\left[-\frac{a_0-a_1}{\sigma}\sqrt{n} + x_\alpha\right].$$

Considérons maintenant la *stratégie du minimax* posant $\Pi_{00} = \Pi_{11} = 0$, $\Pi_{01} = \lambda\Pi_{10}$. On a alors $c^* = \lambda$, $c = \lambda\mu = \frac{q\lambda}{p}$ et l'équation donnant la valeur la moins favorable de la probabilité a priori $q_{\text{mm}} = 1 - p_{\text{mm}}$ devient en vertu de (1.51), (1.73) et (1.73')

$$1 - F\left[\frac{a_1-a_0}{2\sigma}\sqrt{n} + \frac{\sigma}{\sqrt{n}\,(a_1-a_0)}\ln\frac{q\lambda}{1-q}\right] = \\ = \frac{1}{\lambda} F\left[-\frac{a_1-a_0}{2\sigma}\sqrt{n} + \frac{\sigma}{\sqrt{n}\,(a_1-a_0)}\ln\frac{q\lambda}{1-q}\right]. \quad (1.78)$$

Pour $\lambda = 1$, en vertu de (1.78) on a $q_{\text{mm}} = \frac{1}{2}$. Comme on pouvait s'y attendre, lorsque les erreurs de première et de seconde espèce conduisent à des pertes égales, l'apparition équiprobable des états s_0 et s_1 est la moins favorable du point de vue du risque de Bayes.

On a représenté sur la figure 1.5 le risque de Bayes R^* en fonction de q pour $\frac{a_1-a_0}{2\sigma}\sqrt{n} = 1$. Dans ce cas en vertu de (1.23), (1.73) et (1.73') l'expression de R^* est

$$R^* = \Pi_{10}\left\{\lambda q\left[1 - F\left(\ln\frac{q\lambda}{1-q} + 1\right)\right] + pF\left(\ln\frac{q\lambda}{1-q} - 1\right)\right\}.$$

Pour un λ donné, la moins favorable des valeurs de q_{mm} correspond à la valeur maximale de $R_{mm} = R^*(q_{mm})$. Comme on peut le voir sur la figure 1.5, pour $\lambda = 1$ les pertes ne diminuent pas beaucoup pour $q \neq q_{mm}$, tandis que pour $\lambda = 10$ la stratégie du minimax peut paraître trop prudente. Cependant, cette prudence est une garantie que les pertes ne dépasseront jamais R_{mm}. En effet, si pour $\lambda = 10$ on s'écarte un peu de la valeur la moins favorable $q_{mm} = 0{,}3$ et si l'on adopte la stratégie de Bayes pour $q_1 = 0{,}2$, les pertes moyennes ne diminueront que de 20 %. Si en réalité $q \neq q_1$ et si l'on utilise la stratégie de Bayes pour $q = q_1$, le risque moyen R varie en fonction de q d'une façon linéaire : $R(q) = qr_0(q_1) + (1-q)r_1(q_1)$ (tangente au point $q = q_1$ à la courbe $R^*(q)$ pour $\lambda = 10$ sur la figure 1.5) et pour certaines valeurs de q peut être *sensiblement* supérieur à R_{mm} correspondant à la stratégie du minimax (voir la partie hachurée sur la figure 1.5).

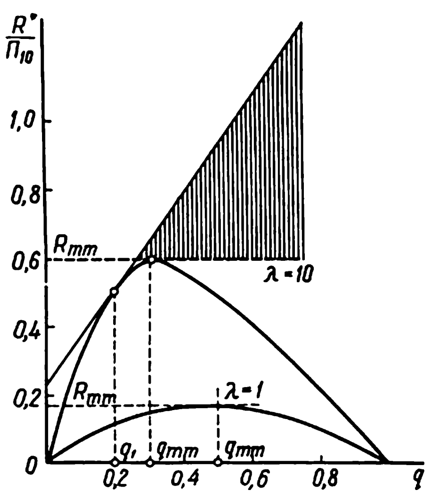


Fig. 1.5. Risque moyen

Considérons en conclusion la stratégie d'*échantillonnage progressif*. Dans ce cas pour des probabilités d'erreur données α et β le logarithme du rapport de vraisemblance (1.66) doit être comparé aux deux seuils $\ln\frac{\beta}{1-\alpha}$ et $\ln\frac{1-\beta}{\alpha}$ (pour $\alpha < 0{,}5$, $\beta < 0{,}5$). Plus exactement, pour $a_1 > a_0$, à la n-ième étape on adopte l'hypothèse d'après laquelle la valeur moyenne de la variable aléatoire est égale à a_0 si

$$\frac{1}{n}\sum_{i=1}^{n} x_i \leqslant \frac{a_0+a_1}{2} + \frac{\sigma^2 \ln\frac{\beta}{1-\alpha}}{n(a_1-a_0)}, \tag{1.79}$$

et l'hypothèse affirmant que la valeur moyenne est égale à a_1 si

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geqslant \frac{a_0+a_1}{2} + \frac{\sigma^2 \ln\frac{1-\beta}{\alpha}}{n(a_1-a_0)}, \tag{1.79'}$$

enfin, on poursuit les observations si ni l'une ni l'autre des inégalités (1.79) et (1.79′) ne se trouve être vérifiée.

Pour calculer les moyennes des tailles minimales des échantillons nécessaires pour aboutir à une décision définitive, il faut en utilisant (1.62) et (1.62′) trouver la moyenne du logarithme du rapport de

vraisemblance unidimensionnel pour les deux cas suivants : 1) lorsque le paramètre de la distribution est égal à a_0 ; 2) lorsque ce paramètre est égal à a_1. Comme en vertu de (1.66)

$$\ln l(x) = \frac{a_1 - a_0}{\sigma^2} x - \frac{a_1^2 - a_0^2}{2\sigma^2},$$

où x est une variable aléatoire normale de variance σ^2 et de moyenne a_0 (ou a_1), dans l'exemple envisagé on a

$$m_{10} = \frac{a_1 - a_0}{\sigma^2} a_0 - \frac{a_1^2 - a_0^2}{2\sigma^2} = -\frac{(a_1 - a_0)^2}{2\sigma^2}, \tag{1.80}$$

$$m_{11} = \frac{a_1 - a_0}{\sigma^2} a_1 - \frac{a_1^2 - a_0^2}{2\sigma^2} = \frac{(a_1 - a_0)^2}{2\sigma^2}. \tag{1.81}$$

En portant (1.80) et (1.81) dans (1.62) et (1.62′) on obtient

$$m_1\{n \mid a_0\} = 2\sigma^2 \frac{(1-\alpha)\ln\frac{1-\alpha}{\beta} + \alpha \ln \frac{\alpha}{1-\beta}}{(a_1 - a_0)^2}, \tag{1.82}$$

$$m_1\{n \mid a_1\} = 2\sigma^2 \frac{\beta \ln\frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha}}{(a_1 - a_0)^2}. \tag{1.83}$$

Ces formules montrent que la taille moyenne des échantillons suffisante pour prendre une décision définitive (γ_0 ou γ_1) est inversement proportionnelle au carré de la grandeur $\frac{a_1 - a_0}{\sigma}$.

Pour $\alpha = \beta = \varepsilon < 0{,}5$ on a

$$m_1\{n \mid a_0\} = m_1\{n \mid a_1\} = 2\left(\frac{\sigma}{a_1 - a_0}\right)^2 (1 - 2\varepsilon) \ln \frac{1-\varepsilon}{\varepsilon}. \tag{1.84}$$

Dans la table 2 on a rapporté les valeurs moyennes $\left(\text{pour } \frac{a_1 - a_0}{\sigma} = 1\right)$ de la taille de l'échantillon calculées d'après (1.84) et la taille de l'échantillon nécessaire pour obtenir les probabilités conditionnelles données $\alpha = \beta = \varepsilon$ pour les stratégies d'échantillonnage de taille fixe suivant le critère du maximum de vraisemblance [cf. (1.74′)]. L'avant-dernière ligne de cette table donne la grandeur du gain lorsque l'on utilise une stratégie d'échantillonnage progressif *)

$$\frac{n}{m_1\{n\}} = \frac{2x_\varepsilon^2}{(1 - 2\varepsilon) \ln \frac{1-\varepsilon}{\varepsilon}}, \tag{1.84′}$$

ce rapport ne dépendant pas de $\frac{\sigma}{a_1 - a_0}$.

*) Ici x_ε est le quantile d'une variable aléatoire normale. Remarquons également que pour le cas $\alpha = \beta$ étudié à titre d'exemple l'échantillonnage progressif n'est pas le plus efficace. Le gain est beaucoup plus important lorsque $\alpha \ll \beta$ ou $\beta \ll \alpha$ (cf. [1], page 57).

Table 2

ε	0,001	0,01	0,05	0,1	0,2
$m_1\{n\}$	13,7	9,0	5,3	3,5	1,67
n	38,2	21,6	10,6	6,5	2,8
$\frac{n}{m_1\{n\}}$	2,78	2,41	2,00	1,85	1,68
$(\varepsilon_{tr})_{max}$	0,0023	0,0197	0,096	0,185	0,355

Notons que pour le cas d'échantillonnage progressif tronqué où $n \leqslant n_{max}$ et $c = 1$, il vient de (1.65) et (1.65′) que les limites supérieures des probabilités d'erreur sont les mêmes et égales à

$$(\varepsilon_{tr})_{max} = \varepsilon + F\left(\frac{d}{2} + \frac{1}{d}\ln\frac{1-\varepsilon}{\varepsilon}\right) - F\left(\frac{d}{2}\right), \qquad (1.84'')$$

où

$$d^2 = n_{max}\left(\frac{a_1 - a_0}{\sigma}\right)^2.$$

La dernière ligne de la table 2 donne les grandeurs $(\varepsilon_{tr})_{max}$ pour $\frac{a_1 - a_0}{\sigma} = 1$ et n_{max} coïncidant avec la taille n de l'échantillon dans le cas d'échantillonnage de taille fixe (troisième ligne de la table).

1.3.2. Variance d'une variable aléatoire normale. Soit une variable aléatoire normale de moyenne nulle. Supposons que l'on émette l'hypothèse H_0 que sa variance est égale à σ_0^2 contre l'alternative simple H_1 que sa variance est égale à σ_1^2. Nous supposons toujours que les éléments de l'échantillon d'après lequel on adopte une décision soient indépendants. Ecrivons l'expression du logarithme du rapport de vraisemblance

$$\ln l(x_1, \ldots, x_n) = \sum_{k=1}^{n} \ln\left[\frac{\sigma_0}{\sigma_1}\exp\left\{-\frac{x_k^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\right\}\right] =$$

$$= n\ln\frac{\sigma_0}{\sigma_1} - \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{k=1}^{n} x_k^2. \qquad (1.85)$$

La stratégie (pour le cas d'échantillonnage de taille fixe) se formule comme suit: on adopte la décision γ_1 (la variance est égale à

σ_1^2) si pour $\sigma_1 > \sigma_0$ on a

$$\sum_{k=1}^{n} x_k^2 \geqslant \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2-\sigma_0^2} \ln\left[\left(\frac{\sigma_1}{\sigma_0}\right)^n c\right] = K, \tag{1.86}$$

et l'on adopte la décision γ_0 (la variance est égale à σ_0^2) si c'est l'inégalité inverse de (1.86) qui est vérifiée.

Ainsi, la vérification de l'hypothèse sur la variance d'une variable aléatoire normale se réduit à la comparaison d'une *somme des carrés* de valeurs échantillonnées au seuil

$$K = \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2-\sigma_0^2} \ln\left[\left(\frac{\sigma_1}{\sigma_0}\right)^n c\right] > 0, \quad \sigma_1 > \sigma_0. \tag{1.86'}$$

La surface séparant le domaine admissible du domaine critique de l'espace des échantillons est dans ce cas une hypersphère avec le centre à l'origine des coordonnées, de rayon $\sqrt{K}$.

En utilisant (1.25) et (1.26) on peut trouver l'expression des probabilités conditionnelles d'erreurs. Comme la somme $\sum_{k=1}^{n}\left(\frac{x_k}{\sigma}\right)^2$ des carrés des variables aléatoires normales normées est distribuée suivant une loi du χ^2 à n degrés de liberté (cf. tome I, problème 3.15), les densités de probabilité de la somme des carrés du premier membre de (1.86) sont

$$w_{10}(y) = \frac{1}{2\Gamma\left(\frac{n}{2}\right)\sigma_0^n}\left(\frac{y}{2}\right)^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma_0^2}}, \quad y > 0, \tag{1.87}$$

si c'est l'hypothèse H_0 qui est vraie, et

$$w_{11}(y) = \frac{1}{2\Gamma\left(\frac{n}{2}\right)\sigma_1^n}\left(\frac{y}{2}\right)^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma_1^2}}, \quad y > 0, \tag{1.87'}$$

si c'est l'hypothèse H_1 qui est vraie. On a alors

$$\alpha = P\left\{\sum_{k=1}^{n} x_k^2 \geqslant K \mid H_0\right\} = \int_K^\infty w_{10}(y)\,dy = 1 - \frac{\Gamma\left(\frac{n}{2}, \frac{K}{2\sigma_0^2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \tag{1.88}$$

$$\beta = P\left\{\sum_{k=1}^{n} x_k^2 < K \mid H_1\right\} = \int_0^K w_{11}(y)\,dy = \frac{\Gamma\left(\frac{n}{2}, \frac{K}{2\sigma_1^2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \tag{1.88'}$$

où $\Gamma(u, v)$ est une fonction gamma incomplète.

La figure 1.6 illustre la stratégie écrite dans (1.86). Les courbes en pointillé donnent les densités de probabilité initiales correspondant aux hypothèses H_0 et H_1. Les courbes en trait continu donnent les distributions de la somme des carrés des valeurs échantillonnées pour les mêmes hypothèses. Les aires hachurées sont égales aux probabilités d'erreurs de première et de seconde espèce.

La grandeur K dans (1.88) et (1.88′) pour les trois premiers critères figurant à la table 1 se calcule directement à partir de (1.86′). Pour le critère de Neumann-Pearson pour un α donné il vient de (1.88)

$$K = \sigma_0^2 \chi_\alpha^2, \tag{1.89}$$

où χ_α^2 est le quantile d'une variable aléatoire distribuée suivant une loi du χ^2 avec n degrés de liberté *). Notons que le seuil

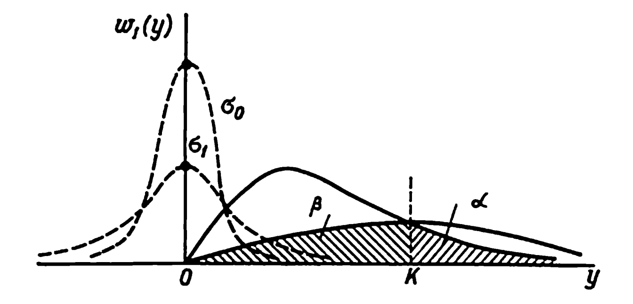


Fig. 1.6. Stratégie de vérification de l'hypothèse sur la variance d'une variable aléatoire normale

K, établi conformément à (1.89), *ne dépend pas* de la grandeur σ_1.

En portant (1.89) dans (1.88′) on obtient la relation

$$\frac{\chi_\alpha^2}{\chi_{1-\beta}^2} = \frac{\sigma_1^2}{\sigma_0^2} \tag{1.89′}$$

entre la probabilité donnée d'erreur de première espèce α, la probabilité minimale possible d'erreur de seconde espèce β et la taille n de l'échantillon, cette dernière est présente dans (1.89′) implicitement, par l'intermédiaire de χ_α^2 et $\chi_{1-\beta}^2$. Il s'ensuit que la taille de l'échantillon avec laquelle on obtient la valeur β (pour α donné) dépend du rapport des variances σ_1^2/σ_0^2.

Si $n \gg 1$ on peut utiliser la relation donnant la tendance asymptotique vers la loi normale d'une variable aléatoire distribuée suivant une loi du χ^2 [cf. tome I, formule (1.43)] et exprimer α et β

*) La table des quantiles de la loi du χ^2 est donnée dans l'*annexe* VIII.

à l'aide de l'intégrale de Laplace [cf. également (1.49) et (1.50)]:

$$\alpha \sim 1 - F\left(\frac{\sqrt{2K}}{\sigma_0} - \sqrt{2n}\right), \tag{1.90}$$

$$\beta \sim F\left(\frac{\sqrt{2K}}{\sigma_1} - \sqrt{2n}\right). \tag{1.90'}$$

Pour $n \to \infty$, en vertu du résultat général mentionné dans le § 1.2.5, les probabilités d'erreurs tendent vers zéro, car si $\sigma_1 > \sigma_0$ on a

$$\frac{\sqrt{2K}}{\sigma_0} - \sqrt{2n} \sim \sqrt{2n}\left[\frac{\frac{\sigma_1}{\sigma_0}\sqrt{\ln\frac{\sigma_1}{\sigma_0}}}{\sqrt{\left(\frac{\sigma_1}{\sigma_0}\right)^2 - 1}} - 1\right] \to \infty,$$

$$\frac{\sqrt{2K}}{\sigma_1} - \sqrt{2n} \sim \sqrt{2n}\left[\frac{\sqrt{\ln\frac{\sigma_1}{\sigma_0}}}{\sqrt{\left(\frac{\sigma_1}{\sigma_0}\right)^2 - 1}} - 1\right] \to -\infty.$$

Remarquons que pour distinguer deux hypothèses voisines $\left(\frac{\sigma_1}{\sigma_0} \to 1\right)$ on doit augmenter la taille de l'échantillon proportionnellement à $\frac{\sigma_0^2}{\sigma_1^2 - \sigma_0^2}$.

Pour la stratégie du minimax, avec $\Pi_{00} = \Pi_{11} = 0$, $\Pi_{01} = \lambda\Pi_{10}$, l'équation donnant la probabilité a priori q_{mm} la moins favorable peut en vertu de (1.51), (1.86'), (1.88) et (1.88') s'écrire de la manière suivante ($\sigma_1 > \sigma_0$):

$$\Gamma\left(\frac{n}{2}\right) - \Gamma\left\{\frac{n}{2}, \frac{\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\ln\left[\left(\frac{\sigma_1}{\sigma_0}\right)^n \frac{q\lambda}{p}\right]\right\} =$$

$$= \frac{1}{\lambda}\Gamma\left\{\frac{n}{2}, \frac{\sigma_0^2}{\sigma_1^2 - \sigma_0^2}\ln\left[\left(\frac{\sigma_1}{\sigma_0}\right)^n \frac{q\lambda}{p}\right]\right\}, \tag{1.91}$$

et pour $n \gg 1$ (en exprimant la fonction gamma incomplète à l'aide de l'intégrale de Laplace) cette équation s'écrit comme suit:

$$\frac{1 - F(c_1\sqrt{2n})}{1 - F(c_2\sqrt{2n})} = \lambda, \tag{1.91'}$$

où

$$c_1 = 1 - \frac{1}{\sqrt{\left(\frac{\sigma_1}{\sigma_0}\right)^2 - 1}}\left[\ln\frac{\sigma_1}{\sigma_0} + \frac{1}{n}\ln\frac{q\lambda}{p}\right]^{1/2};$$

$$c_2 = \frac{\frac{\sigma_1}{\sigma_0}}{\sqrt{\left(\frac{\sigma_1}{\sigma_0}\right)^2 - 1}}\left[\ln\frac{\sigma_1}{\sigma_0} + \frac{1}{n}\ln\frac{q\lambda}{p}\right]^{1/2} - 1. \tag{1.91''}$$

En vertu de (1.91') (à la différence de la stratégie du minimax utilisée pour la vérification de l'hypothèse concernant la moyenne de la taille), pour $\lambda = 1$, la valeur la plus défavorable de q_{mm} est $\frac{1}{2}$. On peut trouver cette grandeur à partir de l'égalité $c_1 = c_2$ (pour $\lambda = 1$) qui, compte tenu de (1.91"), peut s'écrire comme suit:

$$\mu_{mm} = \frac{q_{mm}}{p_{mm}} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[4n \frac{\sigma_1 - \sigma_0}{\sigma_1 + \sigma_0}\right].$$

Passons maintenant à la stratégie d'échantillonnage progressif. Pour des probabilités d'erreurs données α et β et compte tenu de (1.59) elle se formule de la manière suivante : pour $\sigma_1 > \sigma_0$ on accepte l'hypothèse que la variance d'une variable aléatoire normale est égale à σ_0^2, si pour l'échantillon observé on a

$$\sum_{k=1}^{n} x_k^2 \leqslant \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2} \ln\left[\left(\frac{\sigma_1}{\sigma_0}\right)^n \frac{\beta}{1-\alpha}\right], \tag{1.92}$$

et l'hypothèse que la variance est égale à σ_1^2 si on a

$$\sum_{k=1}^{n} x_k^2 \geqslant \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2} \ln\left[\left(\frac{\sigma_1}{\sigma_0}\right)^n \frac{1-\beta}{\alpha}\right]. \tag{1.92'}$$

On continue les observations si ni l'une ni l'autre des inégalités (1.92) et (1.92') ne se trouve vérifiée.

Calculons la valeur moyenne de la taille de l'échantillon nécessaire pour qu'en utilisant (1.62) et (1.62') on puisse prendre une décision définitive. A partir de (1.85) on obtient

$$m_{10} = m_1\{\ln l(x) \mid \sigma_0^2\} =$$
$$= \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2} m_1\{x^2 \mid \sigma_0^2\} + \ln\frac{\sigma_0}{\sigma_1} = \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_1^2} + \ln\frac{\sigma_0}{\sigma_1}, \tag{1.93}$$

$$m_{11} = m_1\{\ln l(x) \mid \sigma_1^2\} =$$
$$= \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2} m_1\{x^2 \mid \sigma_1^2\} + \ln\frac{\sigma_0}{\sigma_1} = \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2} + \ln\frac{\sigma_0}{\sigma_1}. \tag{1.93'}$$

et par conséquent

$$m_1\{n \mid \sigma_0^2\} = 2\,\frac{(1-\alpha)\ln\frac{1-\alpha}{\beta} + \alpha\ln\frac{\alpha}{1-\beta}}{2\ln\frac{\sigma_1}{\sigma_0} - 1 + \frac{\sigma_0^2}{\sigma_1^2}}. \tag{1.94}$$

$$m_1\{n \mid \sigma_1^2\} = 2\,\frac{\beta\ln\frac{\beta}{1-\alpha} + (1-\beta)\ln\frac{1-\beta}{\alpha}}{\frac{\sigma_1^2}{\sigma_0^2} - 1 - 2\ln\frac{\sigma_1}{\sigma_0}}. \tag{1.94'}$$

Les relations (1.94) et (1.94′) montrent que la taille moyenne de l'échantillon nécessaire pour que la décision définitive puisse être prise augmente au fur et à mesure que le rapport $\frac{\sigma_1}{\sigma_0}$ tend vers l'unité.

1.3.3. Paramètre de la loi exponentielle. Soient $x_1, \ldots, x_n$ des valeurs échantillonnées indépendantes distribuées suivant une *loi exponentielle*

$$w_1(x) = \lambda e^{-\lambda x}, \ x > 0, \ \lambda > 0. \tag{1.95}$$

On vérifie l'hypothèse simple H_0 que $\lambda = \lambda_0$ contre l'alternative simple H_1 que $\lambda = \lambda_1 > \lambda_0$. Le logarithme du rapport de vraisemblance est dans ce cas

$$\ln l(x_1, \ldots, x_n) = \sum_{k=1}^{n} \ln \frac{\lambda_1 e^{-\lambda_1 x_k}}{\lambda_0 e^{-\lambda_0 x_k}} = n \ln \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0) \sum_{k=1}^{n} x_k. \tag{1.96}$$

La stratégie (pour une taille n de l'échantillon donnée à l'avance) se formule de la manière suivante: on adopte la décision γ_1 (le paramètre de la distribution est λ_1) si

$$\frac{1}{n} \sum_{k=1}^{n} x_k \leqslant \frac{1}{\lambda_1 - \lambda_0} \ln \frac{\lambda_1}{\lambda_0 c^{1/n}} = K, \tag{1.97}$$

et l'on adopte la décision γ_0 (le paramètre est λ_0) si c'est l'inégalité inverse de (1.97) qui se trouve être vérifiée.

Ainsi, la vérification de l'hypothèse concernant le paramètre de la loi exponentielle revient à la comparaison de la *moyenne arithmétique* des valeurs échantillonnées au seuil

$$K = \frac{1}{\lambda_1 - \lambda_0} \ln \frac{\lambda_1}{\lambda_0 c^{1/n}}, \tag{1.97′}$$

où c dépend du critère de qualité choisi de la stratégie adoptée.

Comme dans le cas de la vérification de l'hypothèse concernant la valeur moyenne d'une variable aléatoire normale, la surface séparant le domaine admissible du domaine critique est un hyperplan perpendiculaire au vecteur unité et écarté de nK de l'origine des coordonnées (K est donné par la formule (1.97′). A la différence de (1.67), le domaine critique (1.97) est disposé au-dessous de l'hyperplan séparateur.

Calculons les probabilités conditionnelles des erreurs de première et de seconde espèce. On sait que la somme de n variables aléatoires indépendantes réparties suivant une loi exponentielle suit une loi du χ^2 avec $2n$ degrés de liberté (cf. tome I, problème 3.16, ainsi que le paragraphe précédent). Compte tenu de la normalisation requise, on trouve que dans l'exemple considéré la variable aléatoire $2\lambda \sum_{k=1}^{n} x_k$

est répartie suivant une loi du χ^2 avec $2n$ degrés de liberté. Ainsi

$$\alpha = P\left\{2\lambda_0 \sum_{k=1}^{n} x_k \leqslant 2n\lambda_0 K \mid H_0\right\} = \frac{\Gamma(n, n\lambda_0 K)}{\Gamma(n)}, \qquad (1.98)$$

$$\beta = P\left\{2\lambda_1 \sum_{k=1}^{n} x_k > 2n\lambda_1 K \mid H_1\right\} = 1 - \frac{\Gamma(n, n\lambda_1 K)}{\Gamma(n)}. \qquad (1.98')$$

La grandeur K dans (1.98) et (1.98'), pour les trois premiers critères figurant dans la table 1, est donnée directement par la formule (1.97'). Pour le critère de Neumann-Pearson, lorsque α est donné, on a en vertu de (1.98) [comparer avec (1.89)]:

$$K = \frac{1}{2n\lambda_0} \chi^2_{1-\alpha}. \qquad (1.99)$$

Ici $\chi^2_{1-\alpha}$ est le quantile de la distribution du χ^2 à $2n$ degrés de liberté. Le seuil calculé d'après (1.99) ne dépend pas de λ_1.

Si $n \gg 1$, compte tenu du fait que la loi du χ^2 tend asymptotiquement vers la loi normale, les relations (1.98) et (1.98') deviennent [cf. également (1.49) et (1.50)]:

$$\alpha \sim F\left(2\sqrt{n\lambda_0 K} - 2\sqrt{n}\right), \qquad (1.100)$$

$$\beta \sim 1 - F\left(2\sqrt{n\lambda_1 K} - 2\sqrt{n}\right). \qquad (1.100')$$

Pour $n \to \infty$, conformément au résultat général donné au § 1.2.5, les probabilités d'erreurs tendent vers zéro, car pour $\lambda_1 > \lambda_0$ on a

$$2\lambda_0 K \sim \frac{1}{\frac{\lambda_1}{\lambda_0} - 1} \ln \frac{\lambda_1}{\lambda_0} < 1,$$

$$2\lambda_1 K \sim \frac{\frac{\lambda_1}{\lambda_0}}{\frac{\lambda_1}{\lambda_0} - 1} \ln \frac{\lambda_1}{\lambda_0} > 1.$$

Pour la stratégie *du minimax*, pour $\Pi_{00} = \Pi_{11} = 0$, $\Pi_{10} = \nu\Pi_{01}$, l'équation donnant la valeur la moins favorable de la probabilité a priori q_{mm} peut en vertu de (1.51), (1.98) et (1.98') s'écrire comme suit:

$$\Gamma(n) - \Gamma\left\{n, \frac{2n\lambda_1}{\lambda_1 - \lambda_0} \ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{\nu p}{q}\right)^{1/n}\right]\right\} =$$

$$= \nu\Gamma\left\{n, \frac{2n\lambda_0}{\lambda_1 - \lambda_0} \ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{\nu p}{q}\right)^{1/n}\right]\right\}, \quad p = 1 - q. \qquad (1.101)$$

Passons maintenant à la stratégie *d'échantillonnage progressif*. Dans l'exemple considéré le domaine intermédiaire G_{int}, pour α et

β donnés, est, en vertu de (1.59) et (1.96), défini par les inégalités

$$\frac{1}{\lambda_1-\lambda_0}\ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{\alpha}{1-\beta}\right)^{1/n}\right]<\frac{1}{n}\sum_{k=1}^{n}x_k<$$

$$<\frac{1}{\lambda_1-\lambda_0}\ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{1-\alpha}{\beta}\right)^{1/n}\right]. \qquad (1.102)$$

Dans le cas où l'inégalité de gauche n'est plus vérifiée, c'est-à-dire si

$$\frac{1}{n}\sum_{k=1}^{n}x_k\leqslant\frac{1}{\lambda_1-\lambda_0}\ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{\alpha}{1-\beta}\right)^{1/n}\right], \qquad (1.102')$$

on adopte la décision γ_1 : le paramètre de la loi de distribution est λ_1 [comparer avec (1.97)], et si c'est l'inégalité de droite qui n'est plus vérifiée, c'est-à-dire si

$$\frac{1}{n}\sum_{k=1}^{n}x_k\geqslant\frac{1}{\lambda_1-\lambda_0}\ln\left[\frac{\lambda_1}{\lambda_0}\left(\frac{1-\alpha}{\beta}\right)^{1/n}\right], \qquad (1.102'')$$

on adopte la décision γ_0 : le paramètre de la loi de distribution est λ_0.

Calculons la valeur moyenne de la taille de l'échantillon nécessaire pour qu'il soit possible de prendre la décision définitive. La moyenne d'une variable aléatoire distribuée suivant une loi exponentielle (1.95) étant égale à $\frac{1}{\lambda}$, on a à partir de (1.63), (1.63')

$$m_{10}=m_1\{\ln l(x)\,|\,\lambda_0\}=$$

$$=\ln\frac{\lambda_1}{\lambda_0}-(\lambda_1-\lambda_0)\frac{1}{\lambda_0}=1-\frac{\lambda_1}{\lambda_0}+\ln\frac{\lambda_1}{\lambda_0}, \qquad (1.103)$$

$$m_{11}=m_1\{\ln l(x)\,|\,\lambda_1\}=$$

$$=\ln\frac{\lambda_1}{\lambda_0}-(\lambda_1-\lambda_0)\frac{1}{\lambda_1}=\frac{\lambda_0}{\lambda_1}-1+\ln\frac{\lambda_1}{\lambda_0}. \qquad (1.103')$$

Par conséquent, pour $\lambda_1>\lambda_0$ on a

$$m_1\{n\,|\,\lambda_0\}=\frac{(1-\alpha)\ln\frac{1-\alpha}{\beta}+\alpha\ln\frac{\alpha}{1-\beta}}{\frac{\lambda_1}{\lambda_0}-1-\ln\frac{\lambda_1}{\lambda_0}}, \qquad (1.104)$$

$$m_1\{n\,|\,\lambda_1\}=\frac{\beta\ln\frac{\beta}{1-\alpha}+(1-\beta)\ln\frac{1-\beta}{\alpha}}{\frac{\lambda_0}{\lambda_1}-1+\ln\frac{\lambda_1}{\lambda_0}}. \qquad (1.104')$$

1.4. HYPOTHÈSES COMPOSITES

1.4.1. Classification des problèmes du choix binaire. Dans le paragraphe précédent nous avons étudié en détail le problème de la vérification d'une hypothèse simple contre une alternative simple. Il s'est trouvé que pour tous les critères de qualité adoptés au début de ce chapitre (§ 1.1.6) (critères de Bayes, du minimax, de Neumann-Pearson, d'échantillonnage progressif) la décision peut s'exprimer par des formules explicites ou des relations asymptotiques simples (vraies pour des échantillons de grande taille). Cependant ce problème ne reflète pas toujours l'état réel des choses. Souvent l'hypothèse vérifiée et l'alternative peuvent être l'une ou l'autre, ou les deux ensemble, *composites*. Dans ce cas on se trouve devant le problème du choix de l'une des deux décisions γ_H ou γ_K, correspondant respectivement à l'authenticité de l'hypothèse H ou de son alternative K, lorsque H et K représentent en général un ensemble d'hypothèses (discrètes ou continues).

Nous nous limitons dans ce paragraphe à l'exposé de la théorie paramétrique où il s'agit de rapporter un échantillon à une classe de distributions donnée par un ou plusieurs *paramètres*. Les valeurs exactes de ces paramètres (ou de certains d'entre eux) sont inconnues. L'information a priori sur un paramètre quelconque s ne fait que nous indiquer que celui-ci appartient à un ensemble de valeurs éventuellement possibles d'un intervalle des paramètres (ou d'un espace de paramètres s'il s'agit de plusieurs paramètres s, u, v, . . .). Ainsi, l'espace des états possibles peut ici être considéré comme l'espace des paramètres d'une classe donnée de distributions des probabilités.

On peut émettre différentes hypothèses concernant les ensembles S_H et S_K des valeurs possibles du paramètre s. Par exemple, une hypothèse composite peut être unilatérale du type $s > s_0$. Les ensembles (intervalles) S_H et S_K peuvent être disjoints ou bien avoir des éléments communs.

1.4.2. Stratégie de Bayes. Considérons la stratégie de Bayes pour le cas général de la vérification d'une hypothèse simple contre une alternative composite. Soient q et $p = 1 - q$ les probabilités a priori que le paramètre inconnu s appartienne respectivement aux ensembles S_H et S_K disjoints, et soient $w_{1H}(s)$ et $w_{1K}(s)$ les densités de probabilité de ce paramètre sur les ensembles mentionnés. Ecrivons la répartition a priori du paramètre s

$$W_1(s) = qw_{1H}(s) + pw_{1K}(s). \tag{1.105}$$

Comme les ensembles S_H et S_K peuvent être discrets, continus ou mixtes, les fonctions $w_{1H}(s)$ et $w_{1K}(s)$ seront respectivement des sommes de fonctions delta, des fonctions continues ou des densités du type mixte [cf. tome I, (2.14), (2.15)]. Par exemple, si l'on vérifie l'hypothèse simple $s = s_0$ contre une alternative composite, dans

le premier terme de (1.105) on a $w_{1H}(s) = \delta(s - s_0)$. (Remarquons que le point $s = s_0$ peut dans ce cas appartenir à l'ensemble K, si seulement la densité de probabilité $w_{1K}(s)$ en ce point est continue.)

Tout comme au § 1.1.2, les décisions possibles γ_0 et γ_1 concernent l'acceptation ou le rejet de l'hypothèse H, c'est pourquoi la matrice des pertes (1.15) se trouve conservée. Cependant, à la différence du § 1.2.2, ici les risques conditionnels [cf. (1.17) et (1.17')] deviennent des *fonctions* du paramètre inconnu s:

$$r_0(s) = \Pi_{00}[1 - \alpha(s)] + \Pi_{01}\alpha(s), \quad (1.106)$$

$$r_1(s) = \Pi_{10}\beta(s) + \Pi_{11}[1 - \beta(s)], \quad (1.106')$$

où

$$\alpha(s) = \int\limits_{G_1}\ldots\int W_n(x_1, \ldots, x_n \mid s)\,dx_1 \ldots dx_n, \quad s \in S_H, \quad (1.107)$$

$$\beta(s) = \int\limits_{G_0}\ldots\int W_n(x_1, \ldots, x_n \mid s)\,dx_1 \ldots dx_n, \quad s \in S_K. \quad (1.107')$$

La moyenne de la fonction de risque peut maintenant s'écrire comme suit [cf. (1.16)]

$$R = q\int\limits_{S_H} r_0(s)\,w_{1H}(s)\,ds + p\int\limits_{S_K} r_1(s)\,w_{1K}(s)\,ds =$$

$$= q\left\{\Pi_{00}\left[1 - \int\limits_{S_H} w_{1H}(s)\,\alpha(s)\,ds\right] + \Pi_{01}\int\limits_{S_H} w_{1H}(s)\,\alpha(s)\,ds\right\} +$$

$$+ p\left\{\Pi_{10}\int\limits_{S_K} w_{1K}(s)\,\beta(s)\,ds + \Pi_{11}\left[1 - \int\limits_{S_K} w_{1K}(s)\,\beta(s)\,ds\right]\right\} =$$

$$= q\Pi_{00} + p\Pi_{10} + q(\Pi_{01} - \Pi_{00})\int\limits_{S_H} w_{1H}(s)\,\alpha(s)\,ds -$$

$$- p(\Pi_{10} - \Pi_{11})\left[1 - \int\limits_{S_K} w_{1K}(s)\,\beta(s)\,ds\right].$$

En remplaçant α et β par leurs expressions (1.107) et (1.107') on obtient

$$R = q\Pi_{00} + p\Pi_{10} - \int\limits_{G_1}\ldots\int\Big[p(\Pi_{10} - \Pi_{11}) \times$$

$$\times \int\limits_{S_K} w_{1K}(s)\,W_n(x_1, \ldots, x_n \mid s)\,ds - q(\Pi_{01} - \Pi_{00}) \times$$

$$\times \int\limits_{S_H} w_{1H}(s)\,W_n(x_1, \ldots, x_n \mid s)\,ds\Big]\,dx_1 \ldots dx_n. \quad (1.108)$$

L'expression (1.108) montre que les mêmes raisonnements que ceux du § 1.2.3 conduisent à la règle optimale suivante basée sur la minimisation du risque moyen: on adopte la décision γ_1 (rejet de l'hypothèse H) si pour l'échantillon observé on a

$$\frac{p \int\limits_{S_K} w_{1K}(s)\, W_n(x_1, \ldots, x_n \mid s)\, ds}{q \int\limits_{S_H} w_{1H}(s)\, W_n(x_1, \ldots, x_n \mid s)\, ds} \geqslant \frac{\Pi_{01} - \Pi_{00}}{\Pi_{10} - \Pi_{11}} = c^*, \qquad (1.109)$$

et l'on adopte la décision γ_0 (acceptation de l'hypothèse H) si c'est l'inégalité contraire de (1.109) qui est vérifiée.

La règle (1.20) est un cas particulier de (1.109) pour

$$w_{1H}(s) = \delta(s - s_0), \quad w_{1K}(s) = \delta(s - s_1).$$

Lors de la vérification d'une hypothèse simple contre une alternative composite il y a lieu d'utiliser le cas particulier de (1.109) pour $w_{1H}(s) = \delta(s - s_0)$. Pour cela on substitue à l'intégrale dans le dénominateur la fonction de vraisemblance $W_n(x_1, \ldots, x_n \mid s_0)$. Les probabilités conditionnelles $\alpha(s)$ et $\beta(s)$ d'erreurs de première et de seconde espèce pour la règle de Bayes (1.109) sont données par les formules (1.25) et (1.26), où F_{10} et F_{11} sont maintenant des fonctions de répartition du rapport des fonctions de vraisemblance moyennes:

$$\Lambda(x_1, \ldots, x_n) = \frac{\int\limits_{S_K} w_{1K}(s)\, W_n(x_1, \ldots, x_n \mid s)\, ds}{\int\limits_{S_H} w_{1H}(s)\, W_n(x_1, \ldots, x_n \mid s)\, ds}, \qquad (1.110)$$

calculées en supposant que le paramètre s de la répartition de l'échantillon $(x_1, \ldots, x_n)$ appartienne à l'ensemble S_H et S_K respectivement.

Si l'hypothèse H est simple on a

$$w_{1H}(s) = \delta(s - s_0)$$

et

$$\Lambda(x_1, \ldots, x_n) = \int\limits_{S_K} w_{1K}(s)\, l(x_1, \ldots, x_n \mid s)\, ds, \qquad (1.110')$$

c'est-à-dire on a la moyenne du rapport de vraisemblance par rapport au paramètre aléatoire.

Les probabilités totales conditionnelles d'erreurs de première et de seconde espèce (pondérées par la densité de probabilité d'apparition de la valeur correspondante du paramètre s) sont

$$\alpha = \int\limits_{S_H} w_{1H}(s)\, \alpha(s)\, ds, \qquad (1.111)$$

$$\beta = \int\limits_{S_K} w_{1K}(s)\, \beta(s)\, ds. \qquad (1.111')$$

Si les domaines S_H et S_K *se recouvrent*, on ne peut plus affirmer avec certitude si une valeur donnée du paramètre appartient au domaine de l'hypothèse ou à celui de l'alternative. Par conséquent, la notion de coût pour une décision correcte ou une décision erronée perd de sens, car la décision peut être correcte ou erronée si s appartient à l'intersection des domaines S_H et S_K. La fonction de pertes faisant correspondre à chaque combinaison de $s \in S_H$, $s \in S_K$ et des décisions γ_0, γ_1 un coût Π_{jk}, dépend maintenant des probabilités a posteriori d'avoir $s \in S_H$ ou $s \in S_K$ si la valeur du paramètre est égale à s (pour des domaines disjoints ces probabilités sont égales à l'unité ou à zéro). On peut raisonner de la manière suivante: supposons que l'on ait $\Pi_{00} = \Pi_{11} = 0$ et que le coût d'une décision erronée soit proportionnel à la probabilité a posteriori pour un certain s de tomber dans le domaine auquel il appartient, c'est-à-dire

$$\Pi_{10}(s) = \Pi_{10} P\{s \in S_K \mid s\} = \Pi(S_K, \gamma_0), \qquad (1.112)$$

$$\Pi_{01}(s) = \Pi_{01} P\{s \in S_H \mid s\} = \Pi(S_H, \gamma_1), \qquad (1.112')$$

où

$$P\{s \in S_K \mid s\} = \frac{P\{s \in S_K\} P\{s \mid s \in S_K\}}{P\{s \in S_K\} P\{s \mid s \in S_K\} + P\{s \in S_H\} P\{s \mid s \in S_H\}} = \frac{p w_{1K}(s)}{p w_{1K}(s) + q w_{1H}(s)}, \qquad (1.113)$$

$$P\{s \in S_H \mid s\} = \frac{P\{s \in S_H\} P\{s \mid s \in S_H\}}{P\{s \in S_K\} P\{s \mid s \in S_K\} + P\{s \in S_H\} P\{s \mid s \in S_H\}} = \frac{q w_{1H}(s)}{p w_{1K}(s) + q w_{1H}(s)}. \qquad (1.113')$$

Si les domaines S_H et S_K sont disjoints on a $P\{s \in S_K \mid s\} = 1$, $P\{s \in S_H \mid s\} = 1$ et $\Pi_{10}(s) = \Pi_{10}$, $\Pi_{01}(s) = \Pi_{01}$, c'est-à-dire qu'on est ramené au cas précédent.

L'expression donnant les risques conditionnels s'écrit maintenant comme suit:

$$r_0(s) = \Pi_{01}(s)\,\alpha(s),$$

$$r_1(s) = \Pi_{10}(s)\,\beta(s),$$

et pour le risque moyen on obtient la formule

$$R = q \int_{S_H} \Pi_{01}(s)\,\alpha(s)\,w_{1H}(s)\,ds + p \int_{S_K} \Pi_{10}(s)\,\beta(s)\,w_{1K}(s)\,ds.$$

1.4.3. Maximum de probabilité a posteriori et maximum de vraisemblance. Les probabilités a posteriori pour le paramètre inconnu s d'appartenir à S_H ou S_K lorsque l'on observe l'échantillon

$(x_1, \ldots, x_n)$ peuvent s'écrire comme suit:

$$P\{s \in S_H \mid x_1, \ldots, x_n\} = \\ = \frac{q \int\limits_{S_H} w_{1H}(s) W_n(x_1, \ldots, x_n \mid s)\, ds}{q \int\limits_{S_H} w_{1H}(s) W_n(x_1, \ldots, x_n \mid s)\, ds + p \int\limits_{S_K} w_{1K}(s) W_n(x_1, \ldots, x_n \mid s)\, ds}. \tag{1.114}$$

$$P\{s \in S_K \mid x_1, \ldots, x_n\} = \\ = \frac{q \int\limits_{S_K} w_{1K}(s) W_n(x_1, \ldots, x_n \mid s)\, ds}{q \int\limits_{S_H} w_{1H}(s) W_n(x_1, \ldots, x_n \mid s)\, ds + p \int\limits_{S_K} w_{1K}(s) W_n(x_1, \ldots, x_n \mid s)\, ds}, \tag{1.114'}$$

La stratégie cherchée est trouvée à partir de la condition que pour rejeter l'hypothèse H il faut que la première des probabilités mentionnées ne soit pas supérieure à la seconde. On a donc la formule

$$\frac{p \int\limits_{S_K} w_{1K}(s) W_n(x_1, \ldots, x_n \mid s)\, ds}{q \int\limits_{S_H} w_{1H}(s) W_n(x_1, \ldots, x_n \mid s)\, ds} \geqslant 1, \tag{1.115}$$

qui est un cas particulier de la règle de Bayes (1.109) pour $\Pi_{01} - \Pi_{00} = \Pi_{10} - \Pi_{11}$.

Le critère de maximum de vraisemblance ne repose pas pour le cas envisagé sur des considérations rigoureuses d'optimalité. On peut éventuellement partir d'une notion purement intuitive du maximum de vraisemblance formulée comme suit: on adopte la décision γ_1 (rejet de l'hypothèse H) si pour l'échantillon observé on a l'inégalité suivante

$$\frac{\max\limits_{s \in S_K} W_n(x_1, \ldots, x_n \mid s)}{\max\limits_{s \in S_H} W_n(x_1, \ldots, x_n \mid s)} \geqslant 1, \tag{1.116}$$

et l'on adopte la décision γ_0 si c'est l'inégalité contraire de (1.116) qui est vérifiée. Remarquons que maintenant la stratégie du maximum de vraisemblance n'est plus un cas particulier de celle de Bayes, comme lors de la vérification d'une hypothèse simple contre une alternative simple.

1.4.4. Critère de Neumann-Pearson. Passons maintenant au critère de Neumann-Pearson. Comme nous l'avons montré au § 1.2.4, quand on vérifie une hypothèse simple contre une alternative simple, il y a toujours, parmi toutes les stratégies pour les

quelles la probabilité de l'erreur de première espèce (niveau de signification) n'est pas supérieure à une valeur donnée α, une stratégie de partition de l'espace des échantillons en domaine critique et domaine admissible qui minimise la probabilité de l'erreur de seconde espèce (cette stratégie a une puissance maximale). Lorsque l'hypothèse H est simple et l'alternative K composite on a $\beta = \beta(s)$ et l'on peut alors essayer de trouver une stratégie de partition de l'espace des échantillons en deux domaines G_0 et G_1 qui, pour une probabilité donnée α de l'erreur de première espèce, minimise la probabilité de l'erreur de seconde espèce $\beta(s)$ [ou maximise la puissance $1 - \beta(s)$] pour *toutes* les alternatives simples constituant l'alternative composite K. Cette stratégie est appelée *uniformément la plus puissante* (on caractérise de même un critère). Si la stratégie uniformément la plus puissante pour la vérification d'une hypothèse simple contre une alternative composite existe, elle ne diffère pratiquement pas de la stratégie analogue correspondant à une alternative *simple*. En effet, la non-univocité due à ce que S_K est un ensemble de valeurs du paramètre s n'a pas d'importance (car le domaine critique ne dépend pas de $s \in S_K$).

L'existence d'une stratégie uniformément la plus puissante est plutôt une exception qu'une règle. Si pour un certain problème cette stratégie n'existe pas, on peut essayer de restreindre la classe des stratégies et de chercher dans cette classe celle qui se trouve être uniformément la plus puissante. Une classe étroite forment des stratégies dites *sans biais*, qui satisfont à la condition suivante : la probabilité de rejeter une hypothèse fausse n'est pas inférieure à la probabilité de rejeter une hypothèse vraie *). Autrement dit, la probabilité α de l'erreur de première espèce se trouve être la limite inférieure des valeurs de la fonction de puissance $1 - \beta(s)$ pour toutes les valeurs de s, c'est-à-dire que l'on a

$$1 - \beta(s) \geqslant \alpha. \tag{1.117}$$

Si $\beta(s)$ est une fonction continue, la valeur minimale de $1 - \beta(s)$ correspond à $s = s_0$ et est précisément égale à α puisque

$$1 - \beta(s_0) = P\{(x_1, \ldots, x_n) \in G_1 \mid s_0\} = \alpha. \tag{1.118}$$

Une stratégie uniformément la plus puissante, si elle existe, est toujours sans biais. Même si elle n'existe pas pour toutes les valeurs du paramètre, on peut quand même trouver une stratégie uniformément la plus puissante pour un certain intervalle de valeurs.

On trouve sur la figure 1.7 les courbes de puissance illustrant la notion de stratégie uniformément la plus puissante, de stratégie uniformément la plus puissante sans biais et de stratégie biaisée

---

*) Notons que dans l'exemple du § 1.1.2 la condition (1.5) est une condition simple de l'absence de biais.

employées pour la vérification d'une hypothèse simple contre une alternative composite.

Remarquons que, basée sur le critère de Neumann-Pearson, la stratégie optimale de vérification d'une hypothèse simple contre une alternative composite n'est pas en général un cas particulier de la règle de Bayes, comme c'était le cas pour une alternative simple.

Les notions étudiées ci-dessus peuvent être étendues au cas où l'hypothèse vérifiée est composite. Dans ce cas la probabilité de

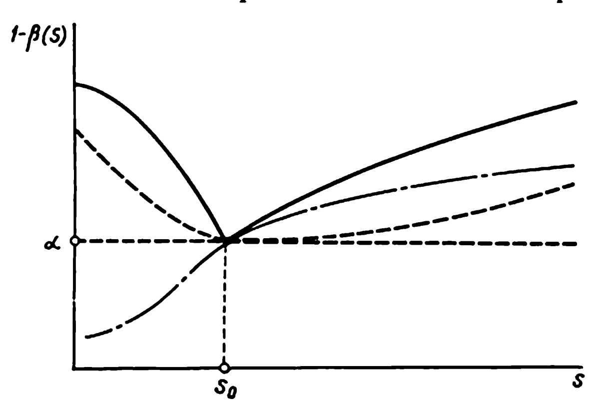


Fig. 1.7. Fonction de puissance:
— stratégie uniformément la plus puissante; --- stratégie uniformément la plus puissante non biaisée; —·—·— stratégie biaisée

l'erreur de première espèce (niveau de signification) dépend également du paramètre s appartenant à un certain ensemble S_H. Cependant on peut parfois indiquer l'ensemble des stratégies (ensemble des domaines critiques) auquel correspond un niveau de signification constant α pour tous les $s \in S_H$. Le problème revient alors à trouver dans l'ensemble des stratégies celle qui sera la plus puissante pour tous les $s \in S_K$. Il est évident que tout comme dans le cas plus simple envisagé ci-dessus, la solution du problème mentionné peut ne pas exister. Il y a alors lieu de la trouver dans une classe plus restreinte de stratégies en introduisant des conditions supplémentaires. L'une de ces conditions peut être l'absence de biais, c'est-à-dire la condition

$$P\{(x_1, \ldots, x_n) \in G_1 \mid s \in S_H\} \leqslant$$
$$\leqslant P\{(x_1, \ldots, x_n) \in G_1 \mid s \in S_K\}. \quad (1.119)$$

1.4.5. Stratégie du minimax. Si l'on connaît les densités $w_{1H}(s)$ et $w_{1K}(s)$ et si l'on ne connaît pas les probabilités q et $p = 1 - q$, on peut prendre la moyenne des fonctions conditionnelles

de risque

$$r_0 = \int\limits_{S_H} r_0(s)\, w_{1H}(s)\, ds, \qquad (1.120)$$

$$r_1 = \int\limits_{S_K} r_1(s)\, w_{1K}(s)\, ds \qquad (1.120)$$

et trouver la stratégie du minimax en égalant les valeurs des risques moyens, tout comme nous l'avons fait au § 1.2.6 dans le cas analogue lors de la vérification d'une hypothèse simple contre une alternative simple.

Dans ce cas, lorsque l'on connaît q mais que l'on ignore $w_{1H}(s)$ et $w_{1K}(s)$, le problème de recherche d'une stratégie du minimax peut ne pas avoir de solution univoque et si de plus on ne connaît pas la probabilité a priori q, il y a lieu d'introduire d'autres critères de qualité.

1.4.6. Echantillonnage progressif. Malheureusement, dans le cas de l'échantillonnage progressif il n'y a pas de situation analogue au cas du critère de Neumann-Pearson lorsqu'une même stratégie pouvait être la meilleure pour tout s appartenant au domaine de l'alternative. La stratégie d'échantillonnage progressif, minimisant la taille moyenne de l'échantillon pour une valeur de s, n'est plus optimale pour une autre valeur du paramètre.

On peut poser le problème de la manière suivante. Vérifier l'hypothèse composite unilatérale H d'avoir $s \leqslant s_0$ contre l'alternative composite d'avoir $s > s_0$. On impose à la stratégie les exigences suivantes:

$$1 - \beta(s) \leqslant \alpha, \qquad s \leqslant s_0, \qquad (1.121)$$

$$1 - \beta(s) \geqslant 1 - \beta, \quad s \geqslant s_1, \; s_1 > s_0. \qquad (1.121')$$

Les conditions (1.121), (1.121′) se trouvent remplies si

$$1 - \beta(s_0) = \alpha, \quad \beta(s_1) = \beta \qquad (1.122)$$

et la probabilité $\beta(s)$ de l'erreur de seconde espèce est une fonction non croissante du paramètre s (fig. 1.8). Il est évident que parmi toutes les stratégies d'échantillonnage progressif satisfaisant aux conditions (1.121) et (1.121′) celle qui consiste à comparer les rapports de vraisemblance aux seuils

$$c_0 = \frac{\beta}{1-\alpha}, \quad c_1 = \frac{1-\beta}{\alpha},$$

minimise la taille moyenne de l'échantillon pour $s = s_0$ et $s = s_1$. En général, la fonction $m_1\{n \mid s\}$ a un maximum entre s_0 et s_1 et décroît lorsque s s'éloigne du point de maximum dans l'un ou l'autre sens. Il peut se faire que ce maximum soit inférieur à la plus

petite des valeurs données de la taille de l'échantillon pour laquelle il existe une stratégie (non progressive) satisfaisant aux conditions (1.121) et (1.121'). Mais comme le montre l'exemple cité dans [7], ceci n'est pas toujours vrai.

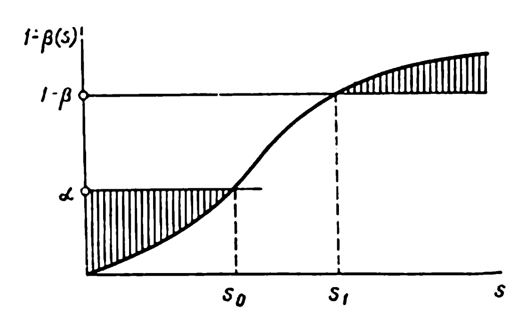


Fig. 1.8. Fonction de puissance

Le calcul exact de la fonction de puissance $1 - \beta(s)$ et de la taille moyenne de l'échantillon $m_1\{n \mid s\}$ dans le cas de l'échantillonnage progressif est très compliqué. Si l'équation transcendante

$$m_1\left\{\left[\frac{w_1(x \mid s_1)}{w_1(x \mid s_0)}\right]^h \middle| s\right\} = 1 \tag{1.123}$$

admet pour solution $h = h(s)$ *), on a approximativement

$$1 - \beta(s) \approx \frac{1 - c_0^{h(s)}}{c_1^{h(s)} - c_0^{h(s)}} \tag{1.124}$$

et

$$m_1\{n \mid s\} \approx \frac{[1 - \beta(s)] \ln c_1 + \beta(s) \ln c_0}{m_1\left\{\ln \frac{w_1(x \mid s_1)}{w_1(x \mid s_0)} \middle| s\right\}} \tag{1.125}$$

en supposant que le dénominateur dans (1.125) soit différent de zéro. Si cette dernière condition n'est pas remplie, au lieu de (1.125) on peut utiliser une autre relation (cf., par exemple, [2], page 223) où dans le dénominateur la moyenne est remplacée par la variance du logarithme du rapport de vraisemblance unidimensionnel:

$$m_1\{n \mid s\} \approx \frac{-\ln c_0 \ln c_1}{m_2\left\{\ln \frac{w_1(x \mid s_1)}{w_1(x \mid s_0)} \middle| s\right\}}. \tag{1.126}$$

1.4.7. Vérification des hypothèses composites concernant la moyenne d'une variable aléatoire normale. Pour illustrer notre exposé, prenons des exemples de vérification des hypothèses sur les paramètres de la loi normale. Commençons par la vérification de l'hypothèse simple H selon laquelle la moyenne d'une variable aléatoire normale est égale à a_0 contre l'alternative composite K, selon laquelle cette moyenne est un nombre réel différent de a_0. On suppose connue la valeur exacte de la variance σ^2 d'une variable aléatoire normale. De plus, on suppose que les éléments de l'échantillon d'après lequel on vérifie l'hypothèse soient indépendants.

*) En vertu de (1.123) on a $h(s_0) = -h(s_1) = 1$.

Pour obtenir la stratégie de Bayes il faut se donner les grandeurs c^*, $\mu = q/p$ ainsi que la distribution a priori de la valeur moyenne lorsque c'est l'alternative K qui est vérifiée. Supposons que cette distribution soit également normale de paramètres (a_1, σ_1^2). On obtient alors à partir de (1.110) pour $w_{1H}(s) = \delta(s - a_0)$ l'expression suivante pour la fonction $\Lambda(x_1, \ldots, x_n)$:

$$\Lambda(x_1, \ldots, x_n) = \exp\left[\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - a_0)^2\right]\frac{1}{\sigma_1\sqrt{2\pi}} \times$$

$$\times \int_{-\infty}^{\infty} e^{-\frac{(s-a_1)^2}{2\sigma_1^2}} e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - s)^2} ds = \left(1 + \frac{n\sigma_1^2}{\sigma^2}\right)^{-\frac{1}{2}} \times$$

$$\times \exp\left\{\frac{1}{2}\left[\frac{2(a_1 - a_0)}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n(a_1^2 - a_0^2)}{\sigma^2} + \right.\right.$$

$$\left.\left. + \frac{\left(\frac{\sigma_1}{\sigma}\right)^2}{1 + n\left(\frac{\sigma_1}{\sigma}\right)^2}\left(\sum_{k=1}^{n}\frac{x_k - a_1}{\sigma}\right)^2\right]\right\},$$

d'où la stratégie de Bayes s'énonce comme suit: on rejette l'hypothèse simple H selon laquelle la moyenne est égale à a_0 si pour $a_1 > a_0$ on a

$$\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{n\left(\frac{\sigma_1}{\sigma}\right)^2}{2\left(1 + \frac{n\sigma_1^2}{\sigma^2}\right)\frac{a_1 - a_0}{\sigma^2}}\left(\frac{1}{n}\sum_{k=1}^{n}\frac{x_k - a_1}{\sigma}\right)^2 \geqslant$$

$$\geqslant \frac{a_0 + a_1}{2} + \frac{\sigma^2}{n(a_1 - a_0)} \ln\left[c\left(1 + \frac{n\sigma_1^2}{\sigma^2}\right)^{1/2}\right] = K, \qquad (1.127)$$

où $c = \mu c^*$. Pour $c = \mu$ on obtient la stratégie correspondant au maximum de probabilité a priori.

Considérons maintenant le critère de Neumann-Pearson. Comme nous l'avons mentionné au § 1.3.1, lors de la vérification de l'hypothèse simple sur la valeur moyenne contre l'alternative simple, la stratégie optimale pour ce critère, déterminée par le seuil (1.76), *ne dépend pas* de l'alternative (à condition que $s > a_0$ ou $s < a_0$). C'est pourquoi la stratégie, selon laquelle on rejette l'hypothèse H si

$$\frac{1}{n}\sum_{k=1}^{n} x_k \geqslant a_0 + \frac{\sigma}{\sqrt{n}} x_\alpha, \qquad (1.128)$$

est uniformément la plus puissante par rapport à l'alternative composite pour laquelle $s > a_0$. Si $s < a_0$, la stratégie uniformément la plus puissante pour une alternative composite détermine le domaine

critique par l'inégalité

$$\frac{1}{n}\sum_{k=1}^{n} x_k \leqslant a_0 - \frac{\sigma}{\sqrt{n}}\, x_\alpha. \tag{1.129}$$

Cependant, si en qualité d'alternative composite on considère toutes les valeurs réelles, la stratégie uniformément la plus puissante n'existe pas. La fonction de puissance pour le critère (1.128) en vertu de (1.77) s'écrit *)

$$1-\beta(s) = 1 - F\left(x_\alpha - \frac{s-a_0}{\sigma}\sqrt{n}\right). \tag{1.130}$$

Pour $s > a_0$, $1-\beta(s) > \alpha$, la fonction de puissance croît d'une façon monotone avec s, et $1-\beta(a_0) = \alpha$. Mais si $s < a_0$ on a

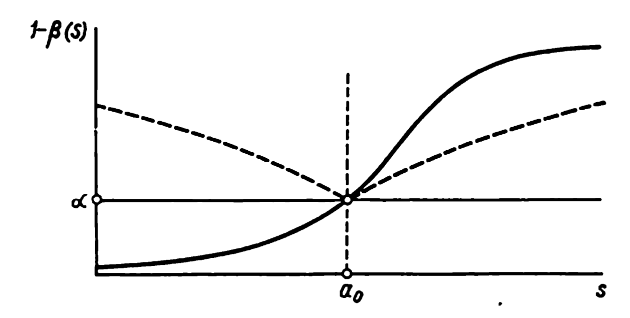


Fig. 1.9. Fonction de puissance pour la vérification de l'hypothèse sur la moyenne d'une distribution : ___ stratégie biaisée; - - - stratégie uniformément la plus puissante non biaisée

$1-\beta(s) < \alpha$ et la fonction de puissance est décroissante lorsque s diminue. Donc le critère (1.128) est *biaisé*.

On peut montrer [6] que la stratégie uniformément la plus puissante est dans ce cas déterminée par le domaine critique

$$\left|\frac{1}{n}\sum_{k=1}^{n} x_i - a_0\right| \geqslant \frac{x_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}. \tag{1.131}$$

La fonction de puissance correspondant au critère (1.131), donnée sur la figure 1.9 en pointillé, est de la forme

$$1-\beta(s) =$$
$$= 1 - \left[F\left(x_{\frac{\alpha}{2}} - \frac{s-a_0}{\sigma}\sqrt{n}\right) - F\left(-x_{\frac{\alpha}{2}} - \frac{s-a_0}{\sigma}\sqrt{n}\right)\right]. \tag{1.132}$$

*) Notons que pour $s = a_1$ la formule (1.130) ne diffère en rien de (1.77).

La fonction (1.132) a un minimum pour $s = a_0$ et

$$1-\beta(a_0)=2-2F(x_{\frac{\alpha}{2}})=\alpha.$$

Pour $s > a_0$ la fonction (1.130) est supérieure à (1.132) car pour un $c > 0$ quelconque on a

$$F(x_{\frac{\alpha}{2}}-c)-F(x_\alpha-c)>F(-x_{\frac{\alpha}{2}}-c).$$

Pour $s = a_0$ les deux fonctions coïncident, mais pour $s < a_0$ la puissance du critère (1.128) est *inférieure* à celle du critère (1.131).

Considérons enfin la stratégie *d'échantillonnage progressif* de vérification de l'hypothèse unilatérale, selon laquelle la valeur moyenne d'une variable aléatoire normale $s \leqslant a_0$, contre l'alternative composite que $s > a_0$. Conformément au § 1.4.6, imposons la condition supplémentaire [cf. (1.121) et (1.121')] selon laquelle pour α et β donnés, on ait

$$1-\beta(s) \leqslant \alpha, \qquad s \leqslant s_0,$$

$$1-\beta(s) \geqslant 1-\beta, \ s \geqslant s_1 > s_0.$$

Pour calculer la fonction de puissance $1-\beta(s)$ il faut résoudre l'équation transcendante (1.123). Le premier membre de cette équation pour l'exemple étudié est égal à

$$m_1\left\{\exp\left[\frac{h}{2\sigma^2}[(x-s_0)^2-(x-s_1)^2]\right]\Big|\,s\right\}=$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(\frac{s_0^2-s_1^2}{2\sigma^2}h\right)\times$$

$$\times\exp\left\{-\frac{1}{2\sigma^2}[(x-s)^2+2(s_0-s_1)hx]\right\}dx=$$

$$=\exp\left\{\frac{s_0^2-s_1^2}{2\sigma^2}h-\frac{s^2}{2\sigma^2}+\frac{1}{2\sigma^2}[s-(s_0-s_1)h]^2\right\}.$$

En vertu de (1.123) il faut égaler cette expression à l'unité ce qui équivaut à annuler l'exposant de l'exponentielle, c'est-à-dire

$$(s_0^2-s_1^2)h-s^2+[s-(s_0-s_1)h]^2=0,$$

d'où

$$h(s)=\frac{s_1+s_0-2s}{s_1-s_0}=1-\frac{2(s-s_0)}{s_1-s_0}. \tag{1.133}$$

La formule (1.124) permet alors de trouver

$$1-\beta(s)\approx\frac{1-c_0^{\,1-2\frac{s-s_0}{s_1-s_0}}}{c_1^{\,1-2\frac{s-s_0}{s_1-s_0}}-c_0^{\,1-2\frac{s-s_0}{s_1-s_0}}}, \tag{1.134}$$

où

$$c_0 = \frac{\beta}{1-\alpha}; \quad c_1 = \frac{1-\beta}{\alpha}.$$

A partir de (1.133) et (1.134) on trouve

$$h(s_0) = 1, \qquad 1-\beta(s_0) = \frac{1-c_0}{c_1-c_0} = \alpha, \tag{1.135}$$

$$h(s_1) = -1, \quad 1-\beta(s_1) = \frac{1-\frac{1}{c_0}}{\frac{1}{c_1}-\frac{1}{c_0}} = 1-\beta. \tag{1.135'}$$

De plus, la fonction $1-\beta(s)$ croît d'une manière monotone

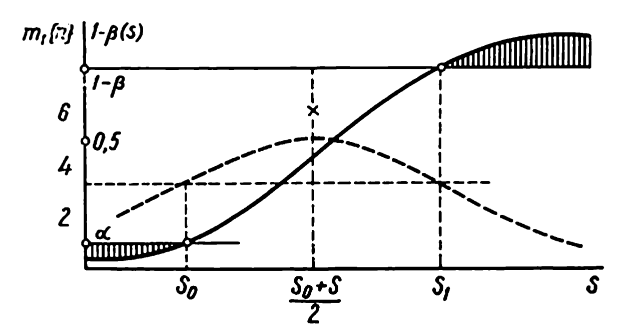


Fig. 1.10. Echantillonnage progressif employé pour la vérification de l'hypothèse sur la moyenne d'une distribution :

—— fonction de puissance; - - - taille moyenne de l'échantillon nécessaire pour qu'une décision puisse être prise

(fig. 1.10) de zéro (pour $s \to -\infty$) à l'unité (pour $s \to \infty$). Pour $s = \frac{s_1+s_0}{2}$ on a

$$1-\beta\left(\frac{s_1+s_0}{2}\right) = \frac{\ln c_0}{\ln \frac{c_0}{c_1}} = \frac{\ln \frac{1-\alpha}{\beta}}{\ln \frac{(1-\alpha)(1-\beta)}{\alpha\beta}}. \tag{1.136}$$

Pour calculer la moyenne conditionnelle, dépendant de s, de la taille de l'échantillon nécessaire pour qu'une décision soit prise, on trouve en vertu de (1.125)

$$m_1\left\{\ln \frac{w_1(x \mid s_1)}{w_1(x \mid s_0)} \middle| s\right\} = m_1\left\{\frac{s_0^2-s_1^2}{2\sigma^2} - \frac{s_0-s_1}{\sigma^2}\, x \middle| s\right\} = \frac{s_1-s_0}{\sigma^2}\left(s - \frac{s_1+s_0}{2}\right)$$

et par conséquent

$$m_1\{n \mid s\} = \frac{[1-\beta(s)]\ln c_1 + \beta(s)\ln c_0}{\frac{s_1-s_0}{\sigma^2}\left(s-\frac{s_1+s_0}{2}\right)}. \tag{1.137}$$

L'expression (1.137) montre que pour $s \to \pm \infty$ on a $m_1 \{n \mid s\} \to 0$ et aux points $s = s_0$ et $s = s_1$ [cf. (1.135), (1.135')]

$$m_1 \{n \mid s_0\} = \frac{\alpha \ln c_1 + (1-\alpha) \ln c_0}{-\frac{(s_1 - s_0)^2}{2\sigma^2}}, \tag{1.138}$$

$$m_1 \{n \mid s_1\} = \frac{(1-\beta) \ln c_1 + \beta \ln c_0}{\frac{(s_1 - s_0)^2}{2\sigma^2}}. \tag{1.139}$$

En fait, ces formules ne diffèrent en rien de (1.82) et (1.83).

Pour $s = \frac{s_1 + s_0}{2}$ le dénominateur de (1.137) s'annule et pour calculer $m_1 \left\{n \,\middle|\, \frac{s_0 + s_1}{2}\right\}$ en ce point, il y a lieu de calculer maintenant le second moment du logarithme du rapport de vraisemblance

$$m_2 \{\ln l(x) \mid s\} = m_1 \left\{\frac{(s_1 - s_0)^2}{\sigma^4} (x - s)^2 \,\middle|\, s\right\} = \frac{(s_1 - s_0)^2}{\sigma^2}. \tag{1.140}$$

En substituant (1.140) dans (1.126) on obtient

$$m_1 \left\{n \,\middle|\, \frac{s_0 + s_1}{2}\right\} = -\frac{\ln c_0 \ln c_1}{\left(\frac{s_1 - s_0}{\sigma}\right)^2}. \tag{1.141}$$

Sur la figure 1.10 la courbe en pointillé donne la taille moyenne de l'échantillon en fonction de s (elle est construite pour $\alpha = \beta = 0{,}1$; voir également la table 2 du § 1.3.1). On voit que la fonction $m_1 \{n \mid s\}$ a un maximum pour $s = \frac{s_0 + s_1}{2}$, celui-ci dépassant à peu près de 40 % la valeur de la taille moyenne minimale de l'échantillon nécessaire pour prendre une décision dans le cas de l'échantillonnage progressif de vérification de l'hypothèse simple $s = s_0$ contre l'alternative simple $s = s_1$. Ce maximum est tout de même inférieur à la taille des échantillons prélevés d'après une stratégie d'échantillonnage fixe pour les mêmes valeurs des probabilités conditionnelles d'erreurs (sur la figure 1.10 cette taille est marquée d'une croix).

1.4.8. Remarque sur les stratégies à plusieurs alternatives. Nous nous limiterons au cas du critère de qualité de Bayes *). Comme nous l'avons noté au § 1.1 pour choisir parmi $m + 1$ hypothèses simples H_j $(j = 0, 1, \ldots, m)$ celle qui est vraie, on peut utiliser le critère de minimum du risque moyen [cf. (1.8)]. La stratégie donne le mode de partition de l'espace des échantillons en $m + 1$ domaines disjoints et attribue (pour une stratégie non randomisée) à chacun des domaines G_k une des décisions γ_k (l'hypothèse H_k suivant laquelle on a l'état s_k de l'effet étudié est vraie).

*) Ces dernières années on a entrepris l'étude des problèmes à plusieurs alternatives. Dans [7] on peut trouver des exemples intéressants de procédures basées sur le critère de Neumann-Pearson et dans [11] les procédures d'échantillonnage progressif, appliquées aux problèmes à plusieurs alternatives.

La probabilité pour que d'après l'échantillon observé $x_1, \ldots, x_n$ on adopte la décision γ_k alors qu'en réalité c'est l'hypothèse H_j qui est vraie sera égale à

$$P\{\gamma_k \mid H_j\} = P\{(x_1, \ldots, x_n) \in G_k \mid s_j\} =$$

$$= \int_{G_k} \cdots \int W_n(x_1, \ldots, x_n \mid s_j)\, dx_1 \ldots dx_n. \qquad (1.142)$$

En portant (1.142) dans (1.8) on obtient l'expression du risque moyen

$$R = \sum_{j=0}^{m} \sum_{k=0}^{m} p_j \Pi_{jk} \int_{G_k} \cdots \int W_n(x_1, \ldots, x_n \mid s_j)\, dx_1 \ldots dx_n, \qquad (1.143)$$

qui dépend du mode de partition de l'espace des échantillons en domaines G_k, $k = 0, 1, \ldots, m$.

Un raisonnement analogue à celui utilisé pour trouver l'inégalité (1.20) permet de démontrer que pour la valeur minimale du risque moyen R l'espace des échantillons G est divisé de telle sorte que le domaine G_k ($k = 1, \ldots, m$) est donné par un *système* de m inégalités, soit

$$\sum_{i=0}^{m} (\Pi_{ij} - \Pi_{ik}) \frac{p_i W_n(x_1, \ldots, x_n \mid s_i)}{p_0 W_n(x_1, \ldots, x_n \mid s_0)} \geqslant 0, \qquad (1.144)$$

$$j = 0, \ldots, m; \; j \neq k.$$

Le domaine G_0 est alors donné par la condition évidente

$$G_0 = G - \sum_{k=1}^{m} G_k. \qquad (1.144')$$

Pour $m = 1$ le système (1.144) contient une seule inégalité coïncidant avec (1.20).

En introduisant les nouvelles variables

$$y_i = \frac{p_i}{p_0} l_i(x_1, \ldots, x_n) = \frac{p_i}{p_0} \frac{W_n(x_1, \ldots, x_n \mid s_i)}{W_n(x_1, \ldots, x_n \mid s_0)},$$

$$i = 1, \ldots, m, \qquad (1.145)$$

c'est-à-dire en représentant les points de l'espace des échantillons dans l'espace à m dimensions des rapports de vraisemblance, on peut écrire le système d'inégalités de la manière suivante:

$$\sum_{i=1}^{m} (\Pi_{ij} - \Pi_{ik}) y_i \geqslant \Pi_{0k} - \Pi_{0j}, \qquad j = 0, \ldots, m; \; j \neq k. \qquad (1.146)$$

Le domaine correspondant au système d'inégalités (1.146) est donné par l'intersection des hyperplans dans l'espace à m dimensions.

Illustrons ce qui vient d'être dit sur l'exemple simple d'un problème à plusieurs alternatives, lorsqu'on sait a priori que trois états s_0, s_1, s_2 sont possibles. En vertu de (1.146) pour $m = 2$ la stratégie de Bayes peut se formuler comme suit :

1) on adopte la décision γ_1 que l'on a l'état s_1 si

$$(\Pi_{10}-\Pi_{11})\, y_1+(\Pi_{20}-\Pi_{21})\, y_2 \geqslant \Pi_{01}-\Pi_{00}, \tag{1.147}$$

$$(\Pi_{12}-\Pi_{11})\, y_1+(\Pi_{22}-\Pi_{21})\, y_2 \geqslant \Pi_{01}-\Pi_{02}; \tag{1.147'}$$

2) on adopte la décision γ_2 que l'on a l'état s_2 si

$$(\Pi_{10}-\Pi_{12})\, y_1+(\Pi_{20}-\Pi_{22})\, y_2 \geqslant \Pi_{02}-\Pi_{00}, \tag{1.148}$$

$$(\Pi_{11}-\Pi_{12})\, y_1+(\Pi_{21}-\Pi_{22})\, y_2 \geqslant \Pi_{02}-\Pi_{01}; \tag{1.148'}$$

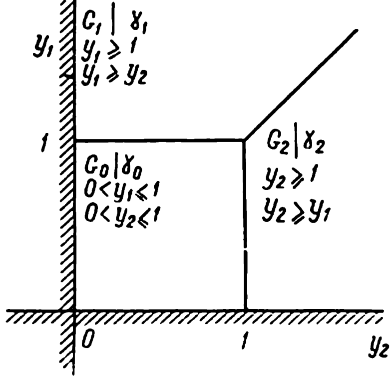


Fig. 1.11. Domaines correspondant à la prise d'une décision

3) on adopte la décision γ_0 que l'on a l'état s_0 si

$$(\Pi_{10}-\Pi_{11})\, y_1+(\Pi_{20}-\Pi_{21})\, y_2 < \Pi_{01}-\Pi_{00}, \tag{1.149}$$

$$(\Pi_{10}-\Pi_{12})\, y_1+(\Pi_{20}-\Pi_{22})\, y_2 < \Pi_{02}-\Pi_{00}. \tag{1.149'}$$

Ici les deux rapports de vraisemblance:

$$y_1=\frac{p_1}{p_0}\,\frac{W_n(x_1, \ldots, x_n \mid s_1)}{W_n(x_1, \ldots, x_n \mid s_0)},$$

$$y_2=\frac{p_2}{p_0}\,\frac{W_n(x_1, \ldots, x_n \mid s_2)}{W_n(x_1, \ldots, x_n \mid s_0)},$$

$$p_0=1-p_1-p_2,$$

donnent la transformation fonctionnelle d'un vecteur aléatoire à n dimensions de composantes $(x_1, \ldots, x_n)$ en un vecteur aléatoire plan de composantes non négatives (y_1, y_2).

Les systèmes ci-dessus de deux inégalités déterminent les domaines disjoints du premier quadrant dans lesquels peut tomber ce vecteur aléatoire plan. Le choix d'une décision est dicté par l'appartenance du vecteur à tel ou tel domaine.

Sur la figure 1.11 on peut voir trois domaines correspondant aux trois décisions possibles pour le cas particulier où les coûts des décisions correctes sont nuls, et les coûts des décisions erronées sont égaux.

1.5. CAS GÉNÉRAUX DU CHOIX D'UNE DÉCISION PARMI DEUX POSSIBLES

1.5.1. Cas de plusieurs paramètres inconnus. Jusqu'à présent nous n'avons envisagé que les problèmes de vérification des hypothèses composites concernant un paramètre d'une distribution contre une alternative composite relative à ce paramètre. Les résultats peuvent être généralisés au cas où la distribution dépend de plusieurs paramètres.

Supposons que la distribution d'une variable aléatoire dépende de M paramètres $s_1, \ldots, s_M$. On vérifie l'hypothèse composite H selon laquelle ledit ensemble de paramètres appartient au domaine S_H de l'espace à M dimensions contre l'alternative composite K que cet ensemble appartient au domaine S_K.

Si l'on connaît les probabilités a priori q et $p = 1 - q$ d'appartenance de l'ensemble $s_1, \ldots, s_M$ aux domaines S_H et S_K disjoints, ainsi que les densités de probabilité conjointes $w_{MH}(s_1, \ldots, s_M)$ et $w_{MK}(s_1, \ldots, s_M)$ de cet ensemble dans les domaines S_H et S_K respectivement et la fonction de pertes, la *stratégie de Bayes* généralisant (1.109) est dans ce cas: on adopte la décision γ_1 (rejet de l'hypothèse H) si pour l'échantillon observé on a

$$\Lambda(x_1, \ldots, x_n) = \frac{\int \ldots \int_{S_K} w_{MK}(s_1, \ldots, s_M) W_n(x_1, \ldots, x_n \mid s_1, \ldots, s_M)\, ds_1 \ldots ds_M}{\int \ldots \int_{S_H} w_{MH}(s_1, \ldots, s_M) W_n(x_1, \ldots, x_n \mid s_1, \ldots, s_M)\, ds_1 \ldots ds_M} \geqslant \frac{q}{p} c^* = c. \qquad (1.150)$$

Les probabilités conditionnelles totales des erreurs de première et de seconde espèce sont

$$\alpha = \int \ldots \int_{S_H} w_{MH}(s_1, \ldots, s_M)\, \alpha(s_1, \ldots, s_M)\, ds_1 \ldots ds_M, \qquad (1.151)$$

$$\beta = \int \ldots \int_{S_K} w_{MH}(s_1, \ldots, s_M)\, \beta(s_1, \ldots, s_M)\, ds_1 \ldots ds_M, \qquad (1.152)$$

où

$$\alpha(s_1, \ldots, s_M) = P\{\Lambda \geqslant c \mid (s_1, \ldots, s_M) \in S_H\}, \qquad (1.153)$$

$$\beta(s_1, \ldots, s_M) = P\{\Lambda < c \mid (s_1, \ldots, s_M) \in S_K). \qquad (1.154)$$

Si l'on ne connaît pas la probabilité a priori p, la valeur la moins favorable de cette probabilité p_{mm} peut être trouvée en égalant les fonctions de risque conditionnelles dont on a pris la moyenne sur les domaines S_H et S_K [cf. § 1.4.5, ainsi que (1.106) et (1.106')]. En portant p_{mm} et $q_{\text{mm}} = 1 - p_{\text{mm}}$ dans le second membre de l'inégalité

(1.150) on obtient la *stratégie du minimax* pour le cas de plusieurs paramètres inconnus.

La recherche d'une stratégie optimale d'après le critère de Neumann-Pearson dans le cas de plusieurs paramètres inconnus est dans le cas général un problème mathématique ardu. A la fin du § 1.4.4 nous n'avons fait que mentionner la possibilité de généralisation des notions de base (l'état actuel de cette question est exposé dans [7]). Tant dans le cas de la vérification d'une hypothèse simple contre une alternative simple, que dans le cas de la vérification d'une *hypothèse simple* contre une *alternative composite*, l'optimisation par le critère de Neumann-Pearson a été obtenue pour une classe de stratégies pour lesquelles la probabilité d'erreur de première espèce (niveau de signification) ne dépassait une valeur donnée α. Dans certains cas de vérification d'une hypothèse composite contre une alternative composite on peut parler d'une classe de stratégies (c'est-à-dire indiquer les domaines critiques G_1) pour lesquelles

$$P\,\{(x_1, \ldots, x_n) \in G_1 \mid (s_1, \ldots, s_M) \in S_H\} = \alpha. \qquad (1.155)$$

Si dans cette classe de stratégies on peut en trouver une qui minimise la probabilité d'erreur de seconde espèce ou a une puissance maximale, c'est-à-dire si

$$1 - \beta = P\,\{(x_1, \ldots, x_n) \in G_1^* \mid (s_1, \ldots, s_M) \in S_K\}, \qquad (1.155')$$

on dit que cette stratégie est uniformément la plus puissante. On peut restreindre la classe des stratégies optimales en introduisant la condition de *l'absence de biais* (1.119).

Nous nous limiterons à un exemple où il y a lieu de vérifier, d'après le critère de Neumann-Pearson, l'hypothèse composite selon laquelle l'échantillon $x_1, \ldots, x_n$ appartient à une loi normale de moyenne a_0 et de *variance inconnue* σ^2, contre l'alternative composite selon laquelle cet échantillon appartient à une loi normale de moyenne $a \neq a_0$ et de variance inconnue. Considérons une variable aléatoire [cf. (2.153)]

$$t = t\,(x_1, \ldots, x_n\,;\ a) =$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_i - a) \left[\frac{1}{n-1} \sum_{h=1}^{n} \left(x_h - \frac{1}{n} \sum_{i=1}^{n} x_i \right)^2 \right]^{-1/2}, \qquad (1.156)$$

qui pour des x_i indépendants et normalement distribués de paramètres (a, σ^2) suit une loi de Student $S_{n-1}(t)$ à $n - 1$ degrés de liberté [cf. (2.154)] ne dépendant pas de a ni de σ^2. Ainsi, les domaines critiques G_1 de l'espace à n dimensions des échantillons auxquels sont rapportés les échantillons $(x_1, \ldots, x_n)$ satisfaisant pour $a = a_0$ à la condition

$$\int_{t_1}^{t_2} S_{n-1}(t)\,dt = \alpha, \qquad (1.156')$$

correspondent aux stratégies mentionnées plus haut pour un niveau de signification constant. On peut montrer que pour $a > a_0$, parmi ces stratégies, uniformément la plus puissante est celle pour laquelle au domaine critique correspondent des valeurs de la variable aléatoire t dépassant le seuil t_α [c'est-à-dire que dans (1.156′) on a $t_2 = \infty$, $t_1 = t_\alpha$].

Autrement dit, on rejette l'hypothèse H selon laquelle la valeur moyenne est égale à a_0 pour une variance inconnue si

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-a_0)\geqslant\left[\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k-\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2\right]^{1/2}\frac{t_\alpha}{\sqrt{n}}, \qquad (1.157)$$

où t_α est le quantile de la loi de Student. Cette stratégie est uniformément la plus puissante par rapport à l'alternative composite $a > a_0$.

En comparant la stratégie (1.157) avec la stratégie analogue (1.128) pour la vérification d'une hypothèse simple relative à la valeur moyenne contre une alternative composite où la variance de la loi normale est connue, on remarque que dans (1.157) la variance inconnue est donnée par l'expression dans les crochets et le quantile x_α de la loi normale a cédé place à celui de la loi de Student (pour α donné).

Avant de calculer la probabilité d'erreur de seconde espèce (ou la puissance $1-\beta$) lorsque l'on utilise la stratégie (1.157), il faut se rendre compte que si les x_i, $i=1, \ldots, n$, suivent une loi normale de paramètre $a \neq a_0$, la variable aléatoire $t(x_1, \ldots, x_n; a_0)$ est distribuée suivant une loi de Student *non centrée*

$$w_1(t;\ \delta)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi(n-1)}}\times$$
$$\times\int_0^\infty y^{\frac{n}{2}-1}\exp\left\{-\frac{1}{2}\left[y+\left(t\sqrt{\frac{y}{n-1}}-\delta\right)^2\right]\right\}dy, \qquad (1.158)$$

où δ est le paramètre de non-centrage égal à

$$\delta=\frac{a-a_0}{\sigma}\sqrt{n}. \qquad (1.158')$$

En utilisant (1.158) on peut écrire l'expression de la fonction de puissance de la stratégie (1.157) sous la forme suivante

$$1-\beta(\delta)=\int_{t_\alpha}^{\infty}w_1(t;\ \delta)\,dt, \qquad (1.159)$$

où t_α est le quantile de la loi de Student non centrée.

Si $a < a_0$ la stratégie uniformément la plus puissante de vérification de l'hypothèse composite selon laquelle la valeur moyenne est égale à a_0 pour une variance inconnue détermine le domaine critique par l'inégalité [cf. (1.129')]

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-a_0)\leqslant-\left[\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k-\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2\right]^{1/2}\frac{t_\alpha}{\sqrt{n}}. \qquad (1.160)$$

Dans le cas où l'alternative contient tous les a réels, il n'existe pas de stratégie uniformément la plus puissante. D'une manière analogue à (1.131), la stratégie suivant laquelle l'hypothèse H est rejetée si

$$\left|\frac{1}{n}\sum_{i=1}^{n}(x_i-a_0)\right|\geqslant\left[\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k-\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2\right]^{1/2}\frac{t_{\alpha/2}}{\sqrt{n}} \qquad (1.161)$$

est la stratégie la plus puissante *non biaisée* avec une probabilité d'erreur de première espèce égale à α.

La probabilité d'erreur de seconde espèce est dans ce cas égale à

$$\beta(\delta)=\int_{-t_{\alpha/2}}^{t_{\alpha/2}} w_1(t;\,\delta)\,dt. \qquad (1.162)$$

Notons en conclusion que la généralisation de la méthode d'échantillonnage progressif au cas de plusieurs paramètres est basée sur des considérations analogues à celles qui ont été exposées au § 1.4.6, avec des complications liées au passage de l'espace unidimensionnel à l'espace multidimensionnel des paramètres.

1.5.2. Echantillon d'une distribution multidimensionnelle. Jusqu'à présent nous avons étudié les problèmes liés à la vérification des hypothèses concernant un ou plusieurs paramètres d'une fonction *unidimensionnelle* de la variable aléatoire ξ d'après un échantillon dont les éléments ont été prélevés sur un ensemble de valeurs possibles de cette variable aléatoire. On peut généraliser le problème au cas de la vérification des hypothèses sur des paramètres de la distribution *multidimensionnelle* d'un ensemble de variables aléatoires $\xi_1, \ldots, \xi_N$. Les éléments d'un échantillon multidimensionnel de taille n seront respectivement des groupes de N' nombres $x_{1k}, x_{2k}, \ldots, x_{Nk}$, $k = 1, \ldots, n$ qui peuvent être considérés comme les composantes d'un vecteur $\mathbf{x}_k$. L'échantillon sera alors une matrice carrée $\mathbf{X}$ de $N \times n$ éléments:

$$\mathbf{X} = \|x_{ik}\|, \; i = 1, \ldots, N; \quad k = 1, \ldots, n. \qquad (1.163)$$

Nous allons illustrer la généralisation de la théorie de vérification des hypothèses concernant les paramètres d'une distribution multidimensionnelle sur l'exemple d'une distribution normale à N di-

mensions. En utilisant les notations vectorielles [cf. (2.57), tome I] on peut écrire la densité de probabilité multidimensionnelle de N variables aléatoires normales liées sous la forme suivante:

$$w_N(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \mathbf{M}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{a})' \mathbf{M}^{-1}(\mathbf{x}-\mathbf{a})\right], \qquad (1.164)$$

où $\mathbf{a}$ est le vecteur colonne des valeurs moyennes; $\mathbf{M}$ est la matrice de corrélation, le signe « prime » indique qu'il s'agit d'une matrice transposée.

Le problème simple de vérification de shypothèses peut ici se formuler comme suit: soit une valeur échantillonnée $\mathbf{x}$ (un vecteur) de la distribution (1.164). On émet l'hypothèse simple H_0 selon laquelle cette valeur suit une loi normale de vecteur des moyennes $\mathbf{a}_0$ et de matrice de corrélation $\mathbf{M}$, contre l'alternative simple H_1 que cette valeur suit une loi normale de vecteur des moyennes $\mathbf{a}_1$ et de même matrice de corrélation $\mathbf{M}$.

Le logarithme du rapport de vraisemblance s'écrit dans ce cas comme suit:

$$\ln l(\mathbf{x}) = \ln \frac{w_N(\mathbf{x} \mid \mathbf{a}_1)}{w_N(\mathbf{x} \mid \mathbf{a}_0)} =$$

$$= -\frac{1}{2}[(\mathbf{x}-\mathbf{a}_1)' \mathbf{M}^{-1}(\mathbf{x}-\mathbf{a}_1) - (\mathbf{x}-\mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{x}-\mathbf{a}_0)] =$$

$$= \mathbf{x}'\mathbf{M}^{-1}(\mathbf{a}_1-\mathbf{a}_0) - \frac{1}{2}(\mathbf{a}_1+\mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{a}_1-\mathbf{a}_0). \qquad (1.165)$$

La structure de cette expression est analogue à celle de (1.66). Les deux formules mentionnées coïncident pour $N = n = 1$.

La stratégie se formule maintenant comme suit: on adopte la décision γ_1 (la moyenne est $\mathbf{a}_1$) si pour le vecteur observé $\mathbf{x}$ on a

$$\mathbf{x}'\mathbf{M}^{-1}(\mathbf{a}_1-\mathbf{a}_0) \geqslant \frac{1}{2}(\mathbf{a}_1+\mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{a}_1-\mathbf{a}_0) + \ln c, \qquad (1.166)$$

et on adopte la décision γ_0 (la moyenne est $\mathbf{a}_0$) si c'est l'inégalité inverse de (1.166) qui se trouve vérifiée. Comme précédemment, la grandeur c dépend du critère choisi (d'après les données a priori dont nous disposons sur l'appartenance à l'une des deux distributions et sur la fonction de pertes).

L'expression (1.166) montre que la procédure de vérification de l'hypothèse concernant la valeur moyenne d'une loi normale multidimensionnelle se réduit au calcul d'une *fonction linéaire* des composantes du vecteur des résultats des observations et à la comparaison au seuil

$$K = \frac{1}{2}(\mathbf{a}_1+\mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{a}_1-\mathbf{a}_0) + \ln c. \qquad (1.167)$$

Pour le calcul des probabilités conditionnelles d'erreurs de première et de seconde espèce, on peut utiliser les formules (1.25) et

(1.26), c'est-à-dire la représentation de l'espace à N dimensions du vecteur $\mathbf{x}$ à l'aide du domaine unidimensionnel du rapport de vraisemblance. En vertu de (1.165), $\ln l(\mathbf{x})$ est une variable aléatoire normale. C'est pourquoi il suffit de calculer sa moyenne et sa variance pour les deux hypothèses H_1 et H_0. On a

$$m_1\{\ln l(\mathbf{x}) \mid H_1\} = m_1\{\mathbf{x}'\} \mathbf{M}^{-1}(\mathbf{a}_1 - \mathbf{a}_0) - \frac{1}{2}(\mathbf{a}_1 + \mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{a}_1 - \mathbf{a}_0) =$$

$$= \mathbf{a}_1' \mathbf{M}^{-1}(\mathbf{a}_1 - \mathbf{a}_0) - \frac{1}{2}(\mathbf{a}_1 + \mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{a}_1 - \mathbf{a}_0) =$$

$$= \frac{1}{2}(\mathbf{a}_1 - \mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{a}_1 - \mathbf{a}_0)$$

et avec les notations

$$d^2 = (\mathbf{a}_1 - \mathbf{a}_0)' \mathbf{M}^{-1}(\mathbf{a}_1 - \mathbf{a}_0), \tag{1.168}$$

on obtient

$$m_1\{\ln l(\mathbf{x}) \mid H_1\} = \frac{d^2}{2}. \tag{1.169}$$

D'une manière analogue on trouve

$$m_1\{\ln l(\mathbf{x}) \mid H_0\} = -\frac{d^2}{2}. \tag{1.170}$$

$$M_2\{\ln l(\mathbf{x}) \mid H_0\} = M_2\{\ln l(\mathbf{x}) \mid H_1\} = d^2. \tag{1.171}$$

Les probabilités d'erreurs de première et de seconde espèce sont alors égales à [cf. (1.73) et (1.73')]

$$\alpha = P\{\ln l(\mathbf{x}) \geqslant \ln c \mid H_0\} = 1 - F\left(\frac{d}{2} + \frac{\ln c}{d}\right), \tag{1.172}$$

$$\beta = P\{\ln l(\mathbf{x}) < \ln c \mid H_1\} = F\left(-\frac{d}{2} + \frac{\ln c}{d}\right). \tag{1.172'}$$

Si les résultats des observations sont représentés non pas par un seul vecteur $\mathbf{x}$ mais par n vecteurs $\mathbf{x}_k$, $k = 1, \ldots, n$, c'est-à-dire par la matrice $\mathbf{X}$ de dimension $N \times n$ [cf. (1.163)], le problème de vérification des hypothèses concernant la valeur moyenne d'une distribution normale multidimensionnelle se réduit au précédent, si l'on remplace $\mathbf{x}$ par la moyenne arithmétique $\frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k$ et la matrice $\mathbf{M}$ par $\frac{\mathbf{M}}{n}$. La grandeur d dans (1.172) et (1.172') doit être remplacée par $d_n = d\sqrt{n}$.

1.5.3. Méthodes non paramétriques de vérification de l'hypothèse de symétrie d'une distribution. Dans les cas non paramétriques, lorsque l'on ne connaît même pas la forme de la distribution à laquelle appartient l'échantillon, la stratégie peut parfois être basée sur l'utilisation d'échantillons ordonnés. Supposons par exemple que l'on vérifie l'hypothèse H concernant la symétrie éventuelle par rapport à l'origine (parité) de la densité de probabilité $w_1(x)$ à laquelle appartient l'échantillon $x_1, \ldots, x_n$.

Introduisons le compteur de signes:

$$u(x)=\begin{cases}1, & x>0,\\ 0, & x\leqslant 0.\end{cases} \qquad (1.173)$$

L'une des stratégies possibles s'énonce comme suit: on rejette l'hypothèse H, c'est-à-dire que l'on adopte la décision γ_1 consistant en ce que

$$w_1(x)\neq w_1(-x) \qquad (1.174)$$

ou

$$F_1(x)\neq 1-F_1(-x), \qquad (1.174')$$

si pour un échantillon $x_1, \ldots, x_n$ de taille donnée n on a

$$\sum_{i=1}^{n} u(x_i)>c, \qquad (1.175)$$

et on adopte la décision γ_0 que la fonction $w_1(x)$ est symétrique par rapport à l'origine si c'est l'inégalité inverse de (1.175) qui se trouve être vérifiée.

Comme les éléments x_i de l'échantillon sont indépendants et ont même distribution, la somme $\sum_{i=1}^{n} u(x_1)$ suit une loi binomiale (cf. § 1.2, tome I) de paramètres n et $p=P\{x_i>0\}=1-F_1(0)$. L'hypothèse H concernant la symétrie de $w_1(x)$ par rapport à l'origine est équivalente à l'affirmation selon laquelle

$$F_1(0)=\frac{1}{2} \qquad (1.176)$$

l'alternative K est alors

$$F_1(0)\neq\frac{1}{2}. \qquad (1.176')$$

Ainsi, la somme dans (1.175) est distribuée suivant une loi binomiale de paramètres $(n, \frac{1}{2})$ pour l'hypothèse H, et suivant la même loi mais de paramètres $\left(n, p\neq\frac{1}{2}\right)$ pour l'alternative K. On peut facilement écrire les expressions des probabilités conditionnelles d'erreurs pour le critère (1.175).

La probabilité conditionnelle d'erreur de première espèce est égale à [cf. (1.22), tome I]

$$\alpha=P\{\gamma_1\mid H\}=P\left\{\sum_{i=1}^{n} u(x_i)>c\mid H\right\}=$$

$$=\sum_{k=[c]+1}^{n}\binom{n}{k}\left(\frac{1}{2}\right)^n=1-I_{\frac{1}{2}}(n-[c],\ [c]+1), \qquad (1.177)$$

où I_q $(n - m, m + 1)$ est le rapport de la fonction bêta incomplète à la fonction bêta complète et $[c]$ la partie entière de c.

La probabilité conditionnelle d'erreur de seconde espèce est égale à

$$\beta = P\{\gamma_0 \mid K\} = P\left\{\sum_{i=1}^{n} u(x_i) \leqslant c \mid K\right\} =$$

$$= \sum_{l=0}^{[c]} \binom{n}{l} p^l (1-p)^{n-l} = I_{1-p}(n-[c], [c]+1). \qquad (1.178)$$

En vertu de (1.177) et (1.178) pour $p = 1 - F_1(0) > \frac{1}{2}$ le critère (1.175) *n'est pas biaisé*. En effet, pour $q = 1 - p < \frac{1}{2}$ en vertu de l'inégalité

$$B_{1/2}(n-m, m+1) = \int_0^{1/2} z^{n-m-1}(1-z)^m dz >$$

$$> \int_0^q z^{n-m-1}(1-z)^m dz = B_q(n-m, m+1)$$

on a l'inégalité

$$I_{1/2}(n-[c]), [c]+1) > I_q(n-[c], [c]+1),$$

et, par conséquent,

$$1 - \beta > \alpha$$

Pour n grand, la loi binomiale tend vers une loi normale (cf. § 1.2.2, tome I) et les formules (1.177) et (1.178) peuvent s'écrire comme suit :

$$\alpha \sim 1 - F\left(\frac{[c]+1-\frac{n}{2}}{\sqrt{n}/2}\right), \qquad (1.179)$$

$$\beta \sim F\left(\frac{[c]+1-np}{\sqrt{np(1-p)}}\right), \qquad (1.180)$$

où $F(x)$ est l'intégrale de Laplace.

Pour un niveau de signification α donné (probabilité d'erreur de première espèce) le seuil c dans le critère (1.175) est donné par la relation suivante

$$[c] = \frac{2x_\alpha}{\sqrt{n}} + \frac{n}{2} - 1, \qquad (1.181)$$

où x_α est le quantile de la loi binomiale (ou asymptotiquement normale pour $n \gg 1$). En portant (1.181) dans (1.180) on obtient

$$\beta \sim F\left[\frac{\frac{2x_\alpha}{n} - \sqrt{n}\left(p - \frac{1}{2}\right)}{\sqrt{p(1-p)}}\right]. \qquad (1.182)$$

Cette relation montre que, pour une stratégie non biaisée $\left(p > \frac{1}{2}\right)$, pour $n \to \infty$, la probabilité d'erreur de seconde espèce $\beta \to 0$.

Si $p = 1 - F_1(0) < \frac{1}{2}$ la stratégie *non biaisée* sera celle pour laquelle l'hypothèse H est rejetée lorsque l'inégalité inverse de (1.175) est vérifiée.

On a alors

$$\alpha = P\left\{\sum_{i=1}^{n} u(x_i) \leqslant c \mid H\right\} = I_{\frac{1}{2}}(n-[c],\ [c]+1), \tag{1.183}$$

$$\beta = P\left\{\sum_{i=1}^{n} u(x_i) > c \mid K\right\} = 1 - I_{1-p}(n-[c],\ [c]+1) \tag{1.183'}$$

et pour $p < \frac{1}{2}$ il vient de (1.183) et (1.183') $1 - \beta > \alpha$. Pour α donné, le seuil c est donné par l'égalité

$$[c] = \frac{2x_{1-\alpha}}{\sqrt{n}} + \frac{n}{2} - 1. \tag{1.184}$$

On peut également formuler une *stratégie bilatérale* en vertu de laquelle l'hypothèse H concernant la symétrie de la distribution est rejetée quand tant le nombre total d'éléments positifs dans l'échantillon que le nombre total d'éléments négatifs dépassent un certain seuil c, c'est-à-dire

$$\sum_{i=1}^{n} u(x_i) > c, \quad n - \sum_{i=1}^{n} u(x_i) > c$$

ou

$$c < \sum_{i=1}^{n} u(x_i) < n - c. \tag{1.185}$$

Les probabilités d'erreurs sont alors égales à

$$\alpha = P\left\{c < \sum_{i=1}^{n} u(x_i) < n - c \mid H\right\} = \sum_{k=[c]+1}^{n-[c]-1} \binom{n}{k} \left(\frac{1}{2}\right)^n =$$

$$= I_{\frac{1}{2}}([c]+1,\ n-[c]) - I_{\frac{1}{2}}(n-[c],\ [c]+1), \tag{1.186}$$

$$\beta = 1 - P\left\{c < \sum_{i=1}^{n} u(x_i) < n - c \mid k\right\} =$$

$$= 1 - \sum_{l=[c]+1}^{n-[c]-1} \binom{n}{l} p^l (1-p)^{n-l} =$$

$$= 1 - I_{1-p}([c]+1,\ n-[c]) + I_{1-p}(n-[c],\ [c]+1). \tag{1.187}$$

Pour $n \gg 1$ les seconds membres de ces formules deviennent des intégrales de Laplace

$$\alpha \sim 1 - 2F\left(\frac{[c]+1-\frac{n}{2}}{\sqrt{n/2}}\right), \tag{1.188}$$

$$\beta \sim F\left(\frac{[c]+1-np}{\sqrt{np(1-p)}}\right) + F\left(\frac{[c]+1-n(1-p)}{\sqrt{np(1-p)}}\right). \tag{1.189}$$

Fixant *) le niveau de signification α on trouve à partir de (1.188) la grandeur du seuil c dans (1.175):

$$c = \frac{2}{\sqrt{n}}\, x_{\frac{1+\alpha}{2}} + \frac{n}{2} - 1, \tag{1.190}$$

où $x_{\frac{1+\alpha}{2}}$ est le quantile de la loi normale.

En portant (1.190) dans (1.189) on obtient

$$\beta \sim F\left[\frac{\frac{2}{\sqrt{n}}\, x_{\frac{1+\alpha}{2}} + \frac{n}{2}(1-2p)}{\sqrt{np(1-p)}}\right] + F\left[\frac{\frac{2}{\sqrt{n}}\, x_{\frac{1+\alpha}{2}} - \frac{n}{2}(1-2p)}{\sqrt{np(1-p)}}\right]. \tag{1.191}$$

Les probabilités d'erreurs peuvent être diminuées si au lieu de (1.175) on utilise un critère plus puissant: l'hypothèse sur la symétrie de la distribution est rejetée si pour l'échantillon $x_1, \ldots, x_n$ de taille donnée n on a

$$\sum_{j=1}^{n} \sum_{i=j}^{n} u(x_i + x_j) > c, \tag{1.192}$$

où $u(x)$ est une fonction donnée par (1.173) (cf. [7], § 6.9).

Problèmes

1.1. Soit H_0 l'hypothèse simple selon laquelle l'échantillon $x_1, \ldots, x_n$ appartient à la distribution exponentielle

$$w_1(x) = e^{-x}, \quad x > 0, \tag{1}$$

et H_1 l'alternative simple selon laquelle cet échantillon appartient à la distribution normale unilatérale

$$w_1(x) = \sqrt{\frac{2}{\pi}}\, e^{-\frac{x^2}{2}}, \quad x > 0. \tag{2}$$

*) Comme $n \gg 1$, on a $[c] \approx c$, et peu importe que le second membre de (1.190) ne soit pas un nombre entier pour α quelconque.

Démontrer que le critère optimal pour rejeter l'hypothèse H_0 est

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-1)^2 \leqslant 1+\ln\frac{2}{\pi}-\frac{2\ln c}{n}, \tag{3}$$

où c est une constante caractérisant le critère de qualité. Montrer que pour $n\to\infty$ on a les formules asymptotiques suivantes pour les probabilités conditionnelles d'erreurs de première et de seconde espèce:

$$\alpha \sim 1-F\left(\frac{\ln c+\frac{n}{2}\ln\frac{\pi}{2}}{\sqrt{2n}}\right), \tag{4}$$

$$\beta \sim F\left[\frac{\ln c-\frac{n}{2}\left(2\sqrt{\frac{2}{\pi}}+\ln\frac{\pi}{2}-1\right)}{\sqrt{\left(\frac{3}{2}-\frac{\pi}{2}-\sqrt{\frac{\pi}{2}}\right)n}}\right]. \tag{4'}$$

1.2. Soient des suites de N expériences indépendantes (cf. § 1.2, tome I). Soit x_i le nombre d'apparitions des événements dans la i-ième suite ($i = 1, \ldots, n$). On émet l'hypothèse simple H_0 que la probabilité a priori d'apparition d'un événement, dans chaque expérience, est égale à p_0, contre l'alternative simple H_1 que cette probabilité est égale à p_1 ($p_1 > p_0$). Démontrer que le critère optimal pour rejeter l'hypothèse H_0 est

$$\frac{1}{n}\sum_{i=1}^{n}x_i \geqslant \frac{\ln\left[\left(\frac{1-p_0}{1-p_1}\right)^N c^{\frac{1}{n}}\right]}{\ln\frac{p_1(1-p_0)}{p_0(1-p_1)}}=K, \tag{5}$$

où c est une constante caractérisant le critère de qualité. Montrer que les probabilités d'erreurs de première et de seconde espèce pour (5) sont respectivement égales à

$$\alpha=1-\frac{B_{1-p_0}(Nn-[nK],\ [nK]+1)}{B(Nn-[nK],\ [nK]+1)}, \tag{6}$$

$$\beta=\frac{B_{1-p_1}(Nn-[nK],\ [nK]+1)}{B(Nn-[nK],\ [nK]+1)}, \tag{6'}$$

où $B_q(x, y)$ et $B(x, y)$ sont respectivement des fonctions bêta complète et incomplète.

1.3. Soit H_0 l'hypothèse simple selon laquelle l'échantillon $x_1, \ldots, x_n$ appartient à la distribution de Rayleigh

$$w_1(x)=\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}} \tag{7}$$

de paramètre $\sigma=\sigma_0$, et H_1, l'alternative simple selon laquelle cet échantillon appartient à la même distribution de paramètre $\sigma=\sigma_1>\sigma_0$. Démontrer que le critère optimal pour rejeter l'hypothèse H_0 est

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 \geqslant \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2-\sigma_0^2}\ln\left(\frac{\sigma_1^2}{\sigma_0^2}c^{\frac{1}{n}}\right)=K, \tag{8}$$

où c est une constante caractérisant le critère de qualité. Montrer que les probabilités d'erreurs de première et de seconde espèce pour (8) sont respectivement

$$\alpha=\frac{\Gamma\left(n,\ \frac{nK}{\sigma_0^2}\right)}{\Gamma(n)}, \tag{9}$$

$$\beta=1-\frac{\Gamma\left(n,\ \frac{nK}{\sigma_1^2}\right)}{\Gamma(n)}, \tag{9'}$$

où $\Gamma(n, x)$ est une fonction gamma incomplète. Expliquer la coïncidence des formules (9) et (9′) avec (1.98) et (1.98′) pour $\sigma^2=\frac{1}{2\lambda}$.

1.4. Montrer que les variances du logarithme du rapport de vraisemblance unidimensionnel pour deux distributions normales de moyennes a_1 et a_0 et de même variance σ^2 sont [cf. (1.47) et (1.48)]

$$M_{20}=M_{21}=\frac{(a_1-a_0)^2}{\sigma^2}. \tag{10}$$

1.5. Soit H_0 l'hypothèse simple selon laquelle x appartient à une distribution normale

$$w_1(x \mid H_0)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}, \tag{11}$$

et H_1 l'alternative simple selon laquelle cette grandeur appartient à une distribution bimodale

$$w_1(x \mid H_1)=\frac{1}{2\sqrt{\pi}}\,[e^{-(x-a)^2}+e^{-(x+a)^2}]. \tag{12}$$

Démontrer que le critère optimal pour rejeter l'hypothèse H_0 est

$$l(x)=\frac{1}{\sqrt{2}}\,e^{a^2}\,[e^{-\frac{(x-2a)^2}{2}}+e^{-\frac{(x+2a)^2}{2}}]\geqslant c. \tag{13}$$

Ne pas oublier que l'intervalle de l'axe réel x déterminé par l'inégalité (13) est biconnexe.

1.6. Soit $\mathbf{x}$ une valeur échantillonnée (vecteur) d'une distribution normale multidimensionnelle de matrice de corrélation $\sigma^2\mathbf{I}$ (où $\mathbf{I}$ est une matrice unité). On émet l'hypothèse H_0 que la moyenne vectorielle est égale à $\mathbf{a}_0$, contre l'alternative H_1 qu'elle est égale à $\mathbf{a}_1$. Montrer que l'espace des échantillons est divisé d'une manière optimale, d'après le critère du maximum de vraisemblance, par un hyperplan perpendiculaire à la ligne joignant les points $\mathbf{x}=\mathbf{a}_0$ et $\mathbf{x}=\mathbf{a}_1$ et divisant cette ligne en deux.

1.7. Soit $\mathbf{x}$ une valeur échantillonnée (vecteur) d'une distribution normale multidimensionnelle. On émet l'hypothèse H_0 que cette valeur appartient à une distribution normale de moyenne vectorielle $\mathbf{a}$ et de matrice de corrélation $\mathbf{M}_0$ contre l'alternative H_1 que cette valeur appartient à une distribution normale de même moyenne et de matrice de corrélation $\mathbf{M}_1$. Montrer que le critère optimal pour adopter la décision γ_1 (la matrice de corrélation est égale à $\mathbf{M}_1$) est

$$(\mathbf{x}-\mathbf{a})'\,(\mathbf{M}_1^{-1}-\mathbf{M}_0^{-1})\,(\mathbf{x}-\mathbf{a})\geqslant 2\ln c+\ln\frac{\det \mathbf{M}_0}{\det \mathbf{M}_1} \tag{14}$$

pour le vecteur $\mathbf{x}$ observé. On adopte la décision γ_0 (la matrice de corrélation est égale à $\mathbf{M}_0$) si c'est l'inégalité contraire de (14) qui est vérifiée.

Montrer qu'avec le changement de variable

$$\mathbf{y} = \mathbf{f}' (\mathbf{x} - \mathbf{a}), \tag{15}$$

où la matrice $\mathbf{f}$ est donnée par

$$\mathbf{M}_1\mathbf{f} = \mathbf{M}_0\mathbf{f}\Lambda \tag{16}$$

et Λ est une matrice diagonale dont les éléments $\lambda_1, \ldots, \lambda_N$ sont les racines de l'équation

$$\det [\mathbf{M}_1 - \lambda\mathbf{M}_0] = 0, \tag{17}$$

on aura pour l'inégalité (14) l'expression suivante:

$$\sum_{i=1}^{N} \left(1 - \frac{1}{\lambda_i}\right) y_i^2 \geqslant 2 \ln c + \sum_{i=1}^{N} \ln \lambda_i. \tag{18}$$

1.8. Supposons que l'on vérifie l'hypothèse simple H que la variance d'une variable aléatoire normale de moyenne nulle est égale à σ_0^2, contre l'alternative composite que cette variance est $\sigma^2 \neq \sigma_0^2$. Démontrer que si le critère pour rejeter l'hypothèse H est

$$\sum_{k=1}^{n} x_k^2 \geqslant \sigma_0^2\chi_\alpha^2, \tag{19}$$

où χ_α^2 est le quantile de la loi du χ^2 à n degrés de liberté; si $\sigma^2 > \sigma_0^2$ il est uniformément le plus puissant vis-à-vis de l'alternative composite.

1.9. Supposons que l'on vérifie l'hypothèse simple H selon laquelle le paramètre λ d'une loi exponentielle est égal à λ_0 contre l'alternative composite que $\lambda \neq \lambda_0$. Démontrer que si le critère pour rejeter l'hypothèse H est

$$\sum_{k=1}^{n} x_k \leqslant \frac{1}{2\lambda_0} \chi_{1-\alpha}^2, \tag{20}$$

où $\chi_{1-\alpha}^2$ est le quantile de la loi du χ^2 à $2n$ degrés de liberté, il est uniformément le plus puissant vis-à-vis de l'alternative composite, si $\lambda > \lambda_0$.

1.10. Démontrer que la fonctionnelle

$$\Phi [\varphi (x)] = \int_{-\infty}^{\infty} w (x) \ln \varphi (x) \, dx \tag{21}$$

avec $\varphi (x) > 0$ et $\int_{-\infty}^{\infty} \varphi (x) \, dx = 1$, admet un maximum pour

$$\varphi (x) = w (x), \tag{22}$$

c'est-à-dire démontrer que

$$\int_{-\infty}^{\infty} w (x) \ln w (x) \, dx \geqslant \int_{-\infty}^{\infty} w (x) \ln \varphi (x) \, dx. \tag{23}$$

Chapitre 2

STATISTIQUE DES VARIABLES ALÉATOIRES

2.1. L'ÉCHANTILLON ET SES CARACTÉRISTIQUES

2.1.1. Fonction de répartition empirique. Revenons au § 1.1.4 et considérons de nouveau une certaine expérience aléatoire et une variable aléatoire, liée à cette expérience en ce sens que les valeurs possibles de cette variable aléatoire sont les résultats des observations. La suite de n observations est caractérisée par un *échantillon* de taille n, dont les éléments sont les valeurs possibles de la variable aléatoire ξ.

Considérons les échantillons obtenus à la suite d'observations indépendantes. On appelle *échantillonnage aléatoire simple* le processus de prélèvement de ces échantillons.

Si l'on connaît la fonction de répartition $F_1(x)$ ou la densité de probabilité $w_1(x)$ de la variable aléatoire ξ, on dit que l'échantillon $x_1, \ldots, x_n$ appartient à la loi $F_1(x)$ ou $w_1(x)$. En arrangeant les éléments de l'échantillon de sorte que leurs valeurs aillent en croissant $x_i \geqslant x_j$ pour $i > j$, on obtient un *échantillon ordonné*. Les éléments de l'échantillon ordonné sont parfois appelés *statistiques d'ordre*. A titre d'exemple de statistiques d'ordre on peut citer la valeur la plus petite x_1, la valeur la plus grande x_n, la médiane de l'échantillon $x_{\frac{n}{2}+1}$ (si n est pair).

Soit $\nu(x)$ le nombre d'éléments de l'échantillon dont la valeur ne dépasse pas un certain seuil x. La fonction en escalier

$$F_1^*(x) = \frac{\nu(x)}{n} = \frac{1}{n} \sum_h u(x - x_h), \tag{2.1}$$

où $u(x)$ est un échelon unité [voir (1.173)], donne la fréquence de l'événement $\xi \leqslant x$ dans la suite de n observations. On appelle cette fréquence *fonction de répartition empirique de l'échantillon*. Cette fonction est l'équivalent statistique de la fonction de répartition d'une variable aléatoire mais ne coïncide pas avec cette dernière (fig. 2.1). La fonction de répartition empirique $F_1^*(x)$ pour $n \to \infty$ *converge en probabilité* vers la fonction de répartition hypothétique $F_1(x)$ si l'échantillon ayant servi à construire la fonction empirique a été obtenu à partir de la distribution $F_1(x)$ (théorème de Glivenko, cf. [3]), c'est-à-dire pour $n \to \infty$ on a en probabilité

$$\sup | F_1(x) - F_1^*(x) | \to 0. \tag{2.2}$$

Lorsque l'échantillon provient d'une distribution continue et que le nombre d'éléments est très grand on peut procéder à un *groupement*. Dans ce cas le domaine des valeurs possibles de la variable

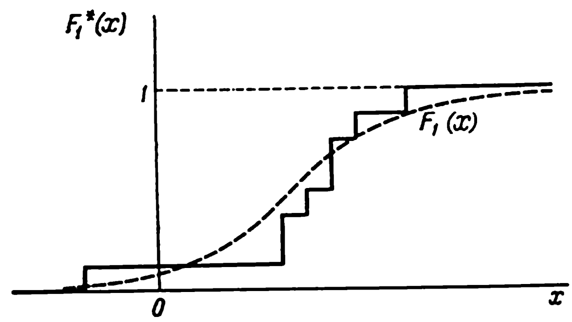


Fig. 2.1. Fonction de répartition empirique

aléatoire est divisé en N intervalles disjoints, puis on rassemble les éléments tombés dans un même intervalle. Dans chacun de ces intervalles pris pour base on construit un rectangle de hauteur $\frac{\nu_i}{nh_i}$ où

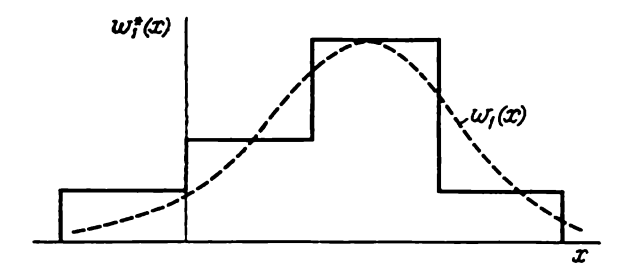


Fig. 2.2. Histogramme

h_i est la longueur de l'intervalle; ν_i le nombre d'éléments tombés dans l'i-ième intervalle; n la taille de l'échantillon. La fonction en escalier ainsi obtenue

$$w_1^*(x)=\frac{1}{n}\sum_{i=1}^{N}\frac{\nu_i(x)}{h_i}, \tag{2.3}$$

$$\nu_i(x)=\begin{cases}\nu_i, & x_i\leqslant x<x_{i+1},\\ 0, & x<x_i,\ x>x_{i+1},\end{cases} \tag{2.3'}$$

$$h_i=x_{i+1}-x_i \tag{2.3''}$$

est appelée *histogramme de l'échantillon*. Elle peut servir d'analogue statistique de la densité de probabilité d'une variable aléatoire (fig. 2.2.)

2.1.2. Moments d'échantillon. La fonction de répartition empirique contient toute l'information statistique que peut fournir un échantillon. La théorie des probabilités nous apprend qu'une variable aléatoire peut être caractérisée non seulement par une fonction de répartition, mais également par des nombres appelés moments de la répartition. D'une manière analogue, les propriétés statistiques d'un échantillon peuvent être caractérisées non seulement par une fonction de répartition empirique, mais également plus grossièrement, par plusieurs nombres appelés *moments d'échantillon.* Le moment d'échantillon d'ordre k est donné par la formule

$$m_k^* = \frac{1}{n} \sum_{i=1}^{n} x_i^k, \tag{2.4}$$

c'est-à-dire qu'il est égal à la moyenne arithmétique des puissances k-ièmes des valeurs échantillonnées. On note d'un astérisque les moments d'échantillon pour ne pas les confondre avec les moments d'une variable aléatoire.

Il faut toujours distinguer les moments d'échantillon des moments d'une variable aléatoire. Ces derniers sont calculés à partir de la densité de probabilité a priori $w_1(x)$ d'après la formule bien connue [cf. (2.77), tome I]

$$m_k = \int_{-\infty}^{\infty} x^k w_1(x)\, dx. \tag{2.5}$$

Pour une variable aléatoire discrète prenant la valeur x_i avec la probabilité p_i, on a

$$m_k = \sum_i x_i^k p_i. \tag{2.5'}$$

Ainsi, les moments d'une variable aléatoire sont calculés d'après les caractéristiques connues a priori, alors que les moments d'échantillon se calculent d'après les valeurs observées.

Etudions les moments d'échantillon des quatre premiers ordres. Par définition, le moment d'échantillon du premier ordre, ou plus simplement la *moyenne échantillonnée*, est égal à la moyenne arithmétique des valeurs échantillonnées, c'est-à-dire

$$m_1^* = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{2.6}$$

La moyenne échantillonnée donne la position de l'échantillon sur l'axe réel, indique l'intervalle où sont groupées les valeurs échantillonnées. La différence $x_i - m_1^*$ est appelée *écart* de la valeur échantillonnée de la moyenne échantillonnée. Les moments d'échantillon de l'écart sont appelés moments centrés et sont désignés par

les symboles M_k^*, $k = 2, 3, \ldots$ En vertu de (2.4) on a

$$M_k^* = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_1^*)^k. \tag{2.7}$$

La *variance échantillonnée* M_2^* (moment d'échantillon centré du second ordre) est la mesure de la dispersion des valeurs échantillonnées par rapport à la moyenne échantillonnée:

$$M_2^* = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_1^*)^2. \tag{2.8}$$

De même, le coefficient d'asymétrie d'échantillon k* et le coefficient d'aplatissement d'échantillon γ^* sont liés aux moments centrés d'échantillon du troisième et du quatrième ordre:

$$\mathrm{k}^* = \frac{M_3^*}{(M_2^*)^{3/2}}, \quad \gamma^* = \frac{M_4^*}{(M_2^*)^2}. \tag{2.9}$$

2.1.3. Fonction de vraisemblance. Chaque échantillon de taille n, c'est-à-dire chaque ensemble de nombres, représente un point dans l'espace à n dimensions. Parmi tous les résultats imaginables des expériences, à l'ensemble de tous les échantillons possibles de taille n correspond un certain domaine de l'espace des échantillons à n dimensions. Nous n'allons pas considérer séparément chaque échantillon, mais l'ensemble de tous les échantillons de taille n appartenant à la distribution a priori $w_1(x)$. La densité de probabilité d'un échantillon à éléments indépendants est égale à [cf. (1.6)]

$$W_n(x_1, \ldots, x_n) = \prod_{k=1}^{n} w_1(x_k). \tag{2.10}$$

La probabilité d'apparition de l'échantillon $x_1, \ldots, x_n$ appartenant à une distribution discrète est

$$P_n(x_1, \ldots, x_n) = \prod_{k=1}^{n} p(x_k), \tag{2.10'}$$

où $p(x_k)$ est la probabilité d'avoir $\xi = x_k$.

La densité de probabilité conjointe $W_n(x_1, \ldots, x_n)$ des valeurs échantillonnées, comme nous l'avons mentionné dans le premier chapitre, est appelée *fonction de vraisemblance de l'échantillon*. Connaissant $w_1(x)$, on peut, à l'aide de la formule (2.10), trouver la fonction de vraisemblance et d'après les règles bien connues de la théorie des probabilités (cf. tome I, chapitre 3) trouver la loi d'une *fonction quelconque* $g(x_1, \ldots, x_n)$ des valeurs échantillonnées.

Les fonctions des valeurs échantillonnées sont souvent appelées *statistiques*. Les moments d'échantillon sont des exemples des statistiques. Les valeurs échantillonnées et leurs fonctions, en parti-

culier, les moments d'échantillon, sont des *variables aléatoires* alors que les caractéristiques numériques (moments) de la distribution initiale $w_1(x)$ sont des nombres constants, bien que certains d'entre eux ou même tous puissent être inconnus *).

Lorsque les données statistiques (échantillons) sont accumulées, on peut utiliser pour la description qualitative des phénomènes étudiés les valeurs échantillonnées ou leurs caractéristiques sous la forme de distributions empiriques, d'histogrammes, de moments d'échantillon, etc. Néanmoins les problèmes de mathématiques statistiques ne sont pas purement descriptifs. L'essentiel est d'obtenir des données statistiques sur une distribution inconnue ou insuffisamment connue d'une variable aléatoire, en se basant sur des valeurs échantillonnées.

2.1.4. Loi des grands nombres. Considérons les échantillons $x_1, \ldots, x_n$ de la distribution $w_1(x)$ de la variable aléatoire ξ, de variance finie σ^2 (et, par conséquent, de moyenne finie a). Soit la moyenne échantillonnée

$$m_1^* = \frac{1}{n}\sum_{i=1}^{n} x_i$$

et calculons sa valeur moyenne sur l'espace des échantillons

$$m_1\{m_1^*\} = \frac{1}{n}\sum_{i=1}^{n} m_1\{x_i\} = \frac{1}{n}\sum_{i=1}^{n} a = a, \tag{2.11}$$

car $m_1\{x_i\} = a$.

Ainsi, la valeur moyenne de la moyenne échantillonnée, quelle que soit la taille de l'échantillon, coïncide exactement avec la moyenne a de la variable aléatoire ξ. Ceci ne veut pas dire évidemment qu'une moyenne échantillonnée quelconque coïncidera avec la grandeur a. La moyenne échantillonnée est une variable aléatoire, mais sa valeur moyenne coïncide exactement avec la moyenne de la distribution initiale.

Calculons maintenant la variance de la moyenne échantillonnée

$$M_2\{m_1^*\} = \frac{1}{n^2}\sum_{i=1}^{n} M_2\{x_i\} = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}. \tag{2.11'}$$

Ici on a utilisé le fait que la variance d'une somme est égale à la somme des variances si les composantes sont indépendantes et si

$$M_2\{x_i\} = \sigma^2.$$

*) Il vient de (2.4) que, compte tenu des restrictions imposées par le théorème central limite (cf. § 3.4, tome I) aux moments de la distribution initiale, les moments d'échantillon m_k^* pour un échantillon de taille importante, sont *asymptotiquement normaux*.

Ainsi, la variance de la moyenne échantillonnée est égale à la variance de la variable aléatoire ξ, divisée par la taille de l'échantillon. Au fur et à mesure que la taille de l'échantillon augmente, la variance de la moyenne échantillonnée diminue et à la limite pour $n \to \infty$ tend vers zéro.

Considérons la probabilité pour la moyenne échantillonnée de s'écarter de la valeur moyenne de la distribution initiale d'une grandeur supérieure à un $\varepsilon > 0$ quelconque. Etudions comment varie cette probabilité lorsque la taille de l'échantillon croît indéfiniment. En vertu de l'inégalité de Tchébychev [cf. (2.92), tome I] cette probabilité ne doit pas être supérieure à la variance de la moyenne échantillonnée divisée par ε^2, c'est-à-dire

$$P\{|m_1^* - a| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{n\varepsilon^2}. \qquad (2.12)$$

Il découle immédiatement de (2.12) que la limite de la probabilité envisagée, lorsque la taille de l'échantillon tend vers l'infini, est nulle:

$$\lim_{n\to\infty} P\{|m_1^* - a| \geqslant \varepsilon\} = 0. \qquad (2.13)$$

La limite de l'inégalité inverse est égale à l'unité

$$\lim_{n\to\infty} P\{|m_1^* - a| < \varepsilon\} = 1. \qquad (2.13')$$

Les formules (2.13) et (2.13′) sont l'expression analytique de la *loi des grands nombres*: au fur et à mesure que la taille de l'échantillon augmente, la probabibité pour que la moyenne échantillonnée diffère aussi peu que l'on veut de la valeur moyenne de la variable aléatoire ξ tend vers l'unité.

Par conséquent, la loi des grands nombres peut se formuler comme suit: la moyenne échantillonnée converge en probabilité vers la moyenne de la distribution initiale (cf. § 3.5, tome I).

De même on peut montrer que tout moment d'échantillon d'ordre k converge en probabilité, lorsque la taille de l'échantillon croît indéfiniment, vers le moment k correspondant de la distribution initiale à condition toutefois qu'il existe un moment d'ordre $2k$ de cette dernière.

En vertu de la loi des grands nombres les caractéristiques échantillonnées peuvent servir d'*estimateurs* des caractéristiques correspondantes de la distribution initiale. Evidemment, on n'obtient une bonne coïncidence des caractéristiques échantillonnées avec les paramètres de la distribution initiale que lorsque les échantillons sont de taille suffisamment grande, ce qui n'est pas toujours facile à réaliser. Au contraire, pour économiser le temps et les moyens on préfère parfois se limiter à des échantillons de taille aussi réduite que possible, mais il est alors difficile d'assurer une bonne précision.

Il est intéressant d'étudier les propriétés de ces estimateurs pour un échantillon de taille finie.

La loi des grands nombres peut s'écrire sous une forme plus générale que (2.13), notamment:

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i)-m_1\{f(x_i)\}\right|\geqslant\varepsilon\right\}=0, \qquad (2.13'')$$

où $f(x)$ est une fonction donnée et $m_2\{f(x_i)\}<\infty$. Si le théorème central limite est vérifié, la distribution $\frac{1}{n}\sum_{i=1}^{n} f(x_i)$ est asymptotiquement normale.

Remarquons en passant qu'on peut utiliser la relation (2.13″) pour donner une estimation statistique de l'intégrale

$$\int_{-\infty}^{\infty} f(x)\,w(x)\,dx=m_1\{f(x)\}\approx\frac{1}{n}\sum_{k=1}^{n} f(x_k),$$

si seulement on dispose d'un échantillon de taille suffisamment importante pris dans la distribution $w(x)$. Cette méthode de calcul des intégrales est appelée *méthode de Monte-Carlo*.

2.2. ESTIMATION DES PARAMÈTRES D'UNE DISTRIBUTION UNIDIMENSIONNELLE

2.2.1. Estimations ponctuelles. De nombreux problèmes d'intérêt pratique reviennent au calcul des données statistiques relatives à une distribution unidimensionnelle dont on connaît la loi mais dont on ignore certains paramètres. On peut procéder par détermination des *estimations ponctuelles* des paramètres, c'est-à-dire qu'on doit calculer les fonctions des valeurs échantillonnées (statistiques)

$$\hat{\vartheta}_n^{(i)}=g^{(i)}(x_1,\ldots,x_n) \qquad (2.14)$$

qui sont les meilleures (en un certain sens) estimations des paramètres inconnus ϑ_i, $i=1,\ldots,m$ de la distribution initiale $w_1(x;\vartheta_1,\ldots,\vartheta_m)$.

Nous allons nous limiter tout d'abord au cas où la distribution en question ne contient qu'un seul paramètre ϑ. Efforçons-nous de donner une classification générale des estimations ponctuelles du paramètre de la distribution unidimensionnelle d'après certaines propriétés de ces estimations. Les plus importantes de ces propriétés sont: la consistance, la suffisance, l'absence de biais et l'efficacité. Considérons séparément chacune de ces propriétés. Notons préalablement que le paramètre estimé ϑ peut lui-même être une variable aléatoire et se caractériser par une distribution a priori $w_1(\vartheta)$.

Cependant, souvent on ne connaît pas cette distribution a priori ou bien on sait d'avance que le paramètre estimé n'est pas aléatoire. En conséquence on a deux cas:

1) une estimation conditionnelle du paramètre ϑ

$$\hat{\vartheta}_n = g(x_1, \ldots, x_n), \tag{2.15}$$

déterminée à partir des valeurs échantillonnées $x_1, \ldots, x_n$ supposant que celles-ci appartiennent à une distribution $w_1(x \mid \vartheta)$ de valeur donnée du paramètre ϑ *);

2) une estimation non conditionnelle du paramètre ϑ

$$\hat{\Theta}_n = \int_{-\infty}^{\infty} w_1(\vartheta)\, g(x_1, \ldots, x_n)\, d\vartheta, \tag{2.15'}$$

obtenue à partir de la moyenne de l'estimation conditionnelle sur toutes les valeurs possibles de ϑ.

Notons que l'estimation conditionnelle aussi bien que l'estimation non conditionnelle, en tant que fonctions des valeurs échantillonnées $x_1, \ldots, x_n$, sont des variables aléatoires dont les distributions sont données par les fonctions de vraisemblance conditionnelle

$$W_n(x_1, \ldots, x_n \mid \vartheta)$$

et inconditionnelle

$$W_n(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} w_1(\vartheta)\, W_n(x_1, \ldots, x_n \mid \vartheta)\, d\vartheta \tag{2.16}$$

de l'échantillon.

2.2.2. Consistance. L'estimation conditionnelle $\hat{\vartheta}_n$ du paramètre ϑ est dite consistante si elle converge en probabilité vers le paramètre estimé lorsque la taille n de l'échantillon augmente indéfiniment, c'est-à-dire pour un $\varepsilon > 0$ arbitraire on a

$$\lim_{n \to \infty} P\{\,|\hat{\vartheta}_n - \vartheta| \geqslant \varepsilon\} = 0. \tag{2.17}$$

D'une manière analogue, l'estimation inconditionnelle $\hat{\Theta}_n$ du paramètre ϑ est dite consistante, si elle converge en probabilité vers la valeur moyenne $m_1\{\vartheta\}$ du paramètre estimé lorsque la taille de l'échantillon augmente indéfiniment, c'est-à-dire si pour un $\varepsilon > 0$ arbitraire on a

$$\lim_{n \to \infty} P\{\,|\hat{\Theta}_n - m_1\{\vartheta\}| \geqslant \varepsilon\} = 0. \tag{2.17'}$$

La loi des grands nombres établit la consistance d'une estimation conditionnelle d'un type particulier, il s'agit de la moyenne échan-

*) On utilise la désignation $w(x \mid \vartheta)$ et des désignations analogues chaque fois où ϑ est un paramètre aléatoire et que l'on étudie la densité conditionnelle de x pour un ϑ donné. Cependant nous utiliserons souvent la désignation $w(x \mid \vartheta)$ tout comme $w(x, \vartheta)$ dans le cas où ϑ est un paramètre inconnu (non aléatoire).

tillonnée convergeant en probabilité vers la moyenne de la distribution initiale. Comme nous l'avons déjà mentionné, on peut généraliser la loi des grands nombres: tout moment d'échantillon converge pour certaines conditions vers le moment correspondant de la distribution initiale. On peut donc affirmer que, pour les conditions mentionnées, tout moment d'échantillon est une estimation consistante du moment correspondant de la distribution initiale.

2.2.3. Absence de biais. L'estimation conditionnelle $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$ du paramètre ϑ est dite non biaisée (absolument correcte), si la valeur moyenne de cette estimation sur l'ensemble des échantillons de taille n donnée est exactement égale au paramètre estimé, c'est-à-dire si pour un n quelconque on a

$$m_1\{\hat{\vartheta}_n\} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) W_n(x_1, \ldots, x_n \mid \vartheta)\, dx_1 \ldots dx_n = \vartheta. \quad (2.18)$$

D'une manière analogue l'estimation inconditionnelle $\hat{\Theta}_n$ du paramètre aléatoire ϑ est dite non biaisée si pour un n quelconque on a *)

$$m_1\{\hat{\Theta}_n\} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \hat{\Theta}_n W_n(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = m_1\{\vartheta\}. \quad (2.19)$$

La différence

$$m_1\{\hat{\vartheta}_n\} - \vartheta = b_n(\vartheta) \quad (2.20)$$

est appelée biais de l'estimation conditionnelle (ou erreur systématique). Pour l'estimation inconditionnelle le biais est donné par

$$m_1\{\hat{\Theta}_n\} - m_1\{\vartheta\} = B_n(\vartheta). \quad (2.21)$$

Notons qu'on peut facilement éliminer le biais $b_n(\vartheta)$ lorsque ce dernier est une fonction linéaire

$$b_n(\vartheta) = a\vartheta + b, \quad (2.22)$$

où a et b sont des nombres réels quelconques (l'un d'eux pouvant être nul). Lorsque ce biais existe réellement, en remplaçant l'estimation $\hat{\vartheta}_n$ par $\frac{\hat{\vartheta}_n - b}{a+1}$ on obtient une estimation non biaisée.

A titre d'exemple d'estimation non biaisée de la moyenne a, pour une distribution arbitraire on peut citer la moyenne échantil-

*) Ici et dans la suite les symboles m_1 au début et à la fin de la formule (2.19) concernent les moyennes prises sur les différents ensembles: pour l'estimation $\hat{\Theta}_n$ la moyenne est prise sur l'espace des échantillons de taille donnée, et celle du paramètre ϑ, sur ses valeurs possibles.

lonnée $\hat{a}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ car

$$m_1\{\hat{a}_n\} = \frac{1}{n}\sum_{i=1}^{n} m_1\{x_i\} = a. \tag{2.23}$$

Ainsi, la moyenne échantillonnée est une estimation consistante et non biaisée de la moyenne de la distribution initiale. Cependant, une estimation consistante n'est pas obligatoirement non biaisée. Ainsi, dans certaines conditions la variance échantillonnée $\hat{\sigma}_n^2 = M_2^*$ est une estimation consistante de la variance σ^2 de la distribution initiale. Cependant

$$m_1\{\hat{\sigma}_n^2\} = \frac{1}{n}\sum_{i=1}^{n} m_1\{(x_i - m_1^*)^2\} =$$

$$= \frac{1}{n}\sum_{i=1}^{n} [m_1\{(x_i - a)^2\} + m_1\{(m_1^* - a)^2\} - 2m_1\{(x_i - a)(m_1^* - a)\}] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\right) = \sigma^2\left(1 - \frac{1}{n}\right), \tag{2.24}$$

c'est-à-dire que l'estimation $\hat{\sigma}_n^2$ est biaisée. La valeur du biais est égale à $-\frac{\sigma^2}{n}$. Lorsque l'échantillon n'est pas grand, le biais peut être important. Par exemple, pour $n = 3$ il atteint 33 %. Comme dans le cas envisagé le biais est une fonction linéaire, en vertu de (2.20), pour $a = -\frac{1}{n}$, $b = 0$, on a pour l'estimation non biaisée de la variance

$$\frac{1}{1 - \frac{1}{n}} M_2^* = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_1^*)^2. \tag{2.25}$$

La consistance de l'estimation ne s'en trouve pas altérée car $\lim_{n\to\infty}\left(1 - \frac{1}{n}\right) = 1$.

Remarquons que le biais de l'estimation $\hat{\sigma}_n^2$ est le coût de l'absence de l'information sur la moyenne a de la distribution initiale. Si a est connue on a

$$m_1\{\hat{\sigma}_n^2\} = \frac{1}{n}\sum_{i=1}^{n} m_1\{(x_i - a)^2\} = \sigma^2$$

et l'estimation mentionnée est absolument correcte.

Parfois on envisage les estimations pour lesquelles les égalités (2.18) et (2.19) ne sont vérifiées que lorsque la taille n de l'échantillon augmente indéfiniment. Les estimations (conditionnelles et inconditionnelles) pour lesquelles

$$\lim_{n\to\infty} m_1 \{\hat{\vartheta}_n\} = \vartheta \tag{2.26}$$

ou

$$\lim_{n\to\infty} m_1 \{\hat{\Theta}_n\} = m_1 \{\vartheta\} \tag{2.27}$$

sont dites *asymptotiquement correctes*. L'estimation (2.24) est un exemple d'estimation asymptotiquement correcte car le biais $b_n(\sigma^2) = -\frac{\sigma^2}{n}$ tend vers zéro pour $n \to \infty$. Il est évident qu'une estimation absolument correcte est toujours asymptotiquement correcte, l'inverse n'étant pas vraie.

2.2.4. Suffisance. Jusqu'à présent on supposait que pour obtenir une estimation il fallait prélever un échantillon de taille donnée n. Cependant, dans certains cas pour l'estimation du paramètre ϑ, il n'est pas indispensable de connaître *chacun* des éléments x_i de l'échantillon, mais il suffit d'avoir une ou plusieurs fonctions $g^{(l)}(x_1, \ldots, x_n)$, $l = 1, \ldots, k$ $(k \leqslant n)$ des valeurs échantillonnées. Ces fonctions sont appelées *estimations suffisantes* (ou *statistiques suffisantes*) du paramètre ϑ. En se limitant aux statistiques suffisantes on peut parfois accélérer notablement le processus d'accumulation des données nécessaires pour estimer un paramètre inconnu.

La condition nécessaire et suffisante pour que $g(x_1, \ldots, x_n)$ soit une statistique suffisante est qu'il soit possible de factoriser la fonction de vraisemblance de l'échantillon, c'est-à-dire de présenter la fonction $W_n(x_1, \ldots, x_n \mid \vartheta)$ comme le produit de deux facteurs non négatifs:

$$W_n(x_1, \ldots, x_n \mid \vartheta) = f[g(x_1, \ldots, x_n), \vartheta]\, h(x_1, \ldots, x_n) \tag{2.28}$$

dont le premier dépend de ϑ et de $g(x_1, \ldots, x_n)$ et le second ne dépend que des valeurs échantillonnées $x_1, \ldots, x_n$ et est indépendant du paramètre estimé ϑ.

Si $g(x_1, \ldots, x_n)$ est une estimation suffisante du paramètre aléatoire ϑ, la distribution a posteriori $W_1(\vartheta \mid x_1, \ldots, x_n)$ dépend non pas des valeurs échantillonnées, mais seulement de la fonction $g(x_1, \ldots, x_n)$. En effet la formule de Bayes [cf. (2.41), tome I] donne

$$W_1(\vartheta \mid x_1, \ldots, x_n) = \frac{w_1(\vartheta)\, W_n(x_1, \ldots, x_n \mid \vartheta)}{\int_{-\infty}^{\infty} w_1(\vartheta)\, W_n(x_1, \ldots, x_n \mid \vartheta)\, d\vartheta}.$$

En y portant l'expression de la fonction de vraisemblance donnée par (2.28) on obtient *)

$$W_1(\vartheta \mid x_1, \ldots, x_n) = \frac{w_1(\vartheta) f[g(x_1, \ldots, x_n), \vartheta]}{\int\limits_{-\infty}^{\infty} w_1(\vartheta) f[g(x_1, \ldots, x_n), \vartheta]\, d\vartheta},$$

c'est-à-dire

$$W_1(\vartheta \mid x_1, \ldots, x_n) = W_1[\vartheta \mid g(x_1, \ldots, x_n)]. \qquad (2.29)$$

Soient $g_1(x_1, \ldots, x_n)$ et $g_2(x_1, \ldots, x_n)$ deux estimations du paramètre ϑ calculées d'après un échantillon de taille donnée n. L'estimation $g_1(x_1, \ldots, x_n)$ est dite subordonnée à l'estimation $g_2(x_1, \ldots, x_n)$ si en vertu de

$$g_2(x_1', \ldots, x_n') = g_2(x_1'', \ldots, x_n'')$$

on a

$$g_1(x_1', \ldots, x_n') = g_1(x_1'', \ldots, x_n'').$$

Toute estimation subordonnant une statistique suffisante est elle-même une estimation suffisante.

Si ϑ_0 est une valeur donnée du paramètre ϑ, on peut montrer (cf. par exemple, [13]) que le *rapport de vraisemblance*

$$l(x_1, \ldots, x_n \mid \vartheta) = \frac{W_n(x_1, \ldots, x_n \mid \vartheta)}{W_n(x_1, \ldots, x_n \mid \vartheta_0)}$$

est une statistique suffisante, permettant de trouver, en utilisant la remarque mentionnée plus haut, une estimation suffisante du paramètre ϑ.

Comme l'a montré E. Dynkine, pour des densités de probabilités unidimensionnelles de la forme

$$w_1(x;\ \vartheta) = \exp\left\{\sum_{i=1}^{r} \varphi_i(x) c_i(\vartheta) + c_0(\vartheta) + \varphi_0(x)\right\},$$

les statistiques du paramètre ϑ existent. La grandeur r est appelée rang de la distribution. Les fonctions $(1, \varphi_1, \ldots, \varphi_r)$ et $(1, c_1, \ldots \ldots, c_r)$ sont linéairement indépendantes.

2.2.5. Efficacité. L'estimation conditionnelle $\hat{\vartheta}_n$ du paramètre ϑ est appelée efficace, si (pour un biais donné) la valeur moyenne (sur l'ensemble des différents échantillons de taille donnée n) du carré de l'écart de l'estimation par rapport au paramètre estimé n'est pas supérieure à l'écart quadratique moyen pour toute autre estimation, c'est-à-dire si

$$m_1\{(\hat{\vartheta}_{n\,\text{eff}} - \vartheta)^2\} \leqslant m_1\{(\hat{\vartheta}_n - \vartheta)^2\}. \qquad (2.30)$$

*) On peut montrer que la quantité d'information $I(x_1, \ldots, x_n, \vartheta)$ contenue dans l'échantillon $x_1, \ldots, x_n$ sur le paramètre inconnu ϑ est égale à la quantité d'information $I[g(x_1, \ldots, x_n), \vartheta]$ contenue dans l'estimation suffisante $g(x_1, \ldots, x_n)$ de ce paramètre (cf. [8], pages 42 à 44).

Autrement dit, l'estimation efficace (si elle existe) minimise pour n donné la grandeur $m_1\{(\hat{\vartheta}_n - \vartheta)^2\}$ sur toutes les estimations conditionnelles $\hat{\vartheta}_n$ (pour un biais donné).

Si $\hat{\vartheta}_n$ est une estimation absolument correcte, on peut dans (2.30) substituer $m_1\{\hat{\vartheta}_n\}$ à ϑ et écrire cette inégalité comme suit:

$$M_2\{\hat{\vartheta}_{n\,\mathrm{eff}}\} \leqslant M_2\{\hat{\vartheta}_n\}. \tag{2.31}$$

Par conséquent, l'estimation efficace non biaisée peut être obtenue à partir de la condition du *minimum de la variance* de l'estimation.

On appelle *efficacité relative* de l'estimation conditionnelle $\hat{\vartheta}_n$ du paramètre ϑ le rapport de l'écart quadratique moyen de l'estimation efficace du paramètre estimé à l'écart quadratique moyen de l'estimation envisagée (pour des estimations biaisées)

$$e_n(\vartheta) = \frac{m_1\{(\hat{\vartheta}_{n\,\mathrm{eff}} - \vartheta)^2\}}{m_1\{(\hat{\vartheta}_n - \vartheta)^2\}} \tag{2.32}$$

et le rapport de la variance de l'estimation efficace à la variance de l'estimation envisagée (pour des estimations non biaisées)

$$e_n(\vartheta) = \frac{M_2\{\hat{\vartheta}_{n\,\mathrm{eff}}\}}{M_2\{\hat{\vartheta}_n\}}. \tag{2.33}$$

Il est évident que $0 \leqslant e_n \leqslant 1$. La grandeur de l'efficacité relative est égale à l'unité pour une estimation efficace.

On parle parfois d'efficacité relative d'une estimation $\hat{\vartheta}_n^{(1)}$ du paramètre ϑ comparativement à une autre estimation $\hat{\vartheta}_n^{(2)}$ de ce même paramètre, en entendant par cela le rapport des écarts quadratiques moyens de ces estimations par rapport au paramètre estimé. Définie de cette façon, l'efficacité relative peut être une grandeur non négative quelconque.

Il y a lieu de distinguer l'estimation efficace de l'estimation asymptotiquement efficace. Une estimation est dite *asymptotiquement efficace*, si la valeur limite de son efficacité relative est différente de zéro, lorsque la taille de l'échantillon augmente indéfiniment. Si l'écart quadratique moyen de l'estimation tend vers zéro pour $n \to \infty$ (ce qui rend l'estimation consistante), l'efficacité signifie que pour un n suffisamment grand les écarts quadratiques moyens par rapport au paramètre estimé de l'estimation considérée et de l'estimation efficace sont des infiniment petits du même ordre de grandeur.

Notons que l'efficacité d'une estimation biaisée est donnée non pas par sa variance, mais par son écart quadratique moyen par rapport au paramètre estimé. On peut facilement donner un exemple simple d'estimation biaisée de variance nulle. Soit $\hat{\vartheta}_n = c_0 = \mathrm{const}$ indépendamment des résultats des mesures. On a alors $M_2\{\hat{\vartheta}_n\} = 0$.

Mais si l'on n'a pas deviné ou bien si l'on ne connaît pas d'avance la grandeur du paramètre estimé, le biais $b_n(\vartheta) = c_0 - \vartheta$ sera grand. En général, il est impossible d'annuler la variance et le biais en même temps. C'est pourquoi dans la suite nous n'allons pas envisager les cas où la variance est nulle (c'est-à-dire quand $b'_n(\vartheta) = -1$).

Il existe une inégalité qui porte le nom de Rao-Cramer, permettant de trouver la limite inférieure des écarts quadratiques moyens des estimations et d'estimer par là même l'efficacité de chacune des estimations.

Supposons que les limites du domaine de l'axe réel x, où la densité de probabilité $w_1(x; \vartheta)$ n'est pas nulle, ne dépendent pas de ϑ. Ceci est vrai, par exemple, si $w_1\{x; \vartheta\} \neq 0$ sur tout l'axe réel ou pour $x > 0$. A titre d'exemple de distribution ne satisfaisant pas à cette condition on peut citer la distribution uniforme

$$w_1(x;\ \vartheta) = \frac{1}{\vartheta}, \qquad 0 < x < \vartheta.$$

Supposons de plus que la fonction $w_1(x; \vartheta)$ soit dérivable par rapport au paramètre ϑ. Soit $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$ une estimation conditionnelle du paramètre ϑ obtenue d'après un échantillon de taille n. La valeur moyenne de cette estimation est

$$m_1\{\hat{\vartheta}_n\} = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} g(x_1, \ldots, x_n) W_n(x_1, \ldots, x_n \mid \vartheta)\, dx_1 \ldots dx_n. \quad (2.34)$$

En utilisant (2.20), on peut écrire (2.34) sous la forme suivante :

$$\vartheta + b_n(\vartheta) = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} g(x_1, \ldots, x_n) W_n(x_1, \ldots, x_n \mid \vartheta)\, dx_1 \ldots dx_n. \quad (2.35)$$

Supposons que l'intégrale dans le second membre de (2.35) soit dérivable par rapport au paramètre ϑ. En dérivant les deux membres de (2.35) par rapport à ϑ et en supposant que les limites d'intégration soient indépendantes de ϑ, on obtient

$$1 + b'_n(\vartheta) = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} g(x_1, \ldots, x_n) \frac{\partial}{\partial\vartheta} W_n(x_1, \ldots, x_n \mid \vartheta) \times$$
$$\times\, dx_1 \ldots dx_n = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} g(x_1, \ldots, x_n) \times$$
$$\times\, W_n(x_1, \ldots, x_n \mid \vartheta) \frac{\partial}{\partial\vartheta} \ln W_n(x_1, \ldots, x_n \mid \vartheta)\, dx_1 \ldots dx_n =$$
$$= m_1\left\{\hat{\vartheta}_n \frac{\partial}{\partial\vartheta} \ln W_n\right\}. \quad (2.36)$$

De plus, en dérivant par rapport à ϑ l'égalité évidente

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} W_n(x_1, \ldots, x_n \mid \vartheta)\, dx_1 \ldots dx_n = 1,$$

on tire

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial}{\partial\vartheta} W_n(x_1, \ldots, x_n \mid \vartheta)\, dx_1 \ldots dx_n =$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} W_n(x_1, \ldots, x_n \mid \vartheta) \frac{\partial}{\partial\vartheta} \ln W_n(x_1, \ldots$$

$$\ldots, x_n \mid \vartheta)\, dx_1 \ldots dx_n = m_1\left\{\frac{\partial}{\partial\vartheta} \ln W_n\right\} = 0. \qquad (2.37)$$

En multipliant l'expression (2.37) par $m_1\{\hat{\vartheta}_n\}$ et en la retranchant de (2.36) on obtient

$$m_1\left\{[\hat{\vartheta}_n - m_1\{\hat{\vartheta}_n\}] \frac{\partial}{\partial\vartheta} \ln W_n\right\} = 1 + b_n'(\vartheta). \qquad (2.38)$$

Le premier membre de (2.38) est la covariance de deux variables aléatoires de moyenne nulle. Le carré de cette covariance ne peut être supérieur au produit des variances des facteurs (cf. § 2.3.5, tome I), c'est-à-dire

$$[1 + b_n'(\vartheta)]^2 \leqslant M_2\{\hat{\vartheta}_n\}\, M_2\left\{\frac{\partial}{\partial\vartheta} \ln W_n\right\}. \qquad (2.39)$$

La grandeur non négative $M_2\left\{\frac{\partial}{\partial\vartheta} \ln W_n\right\} = m_1\left\{\left[\frac{\partial}{\partial\vartheta} \ln W_n\right]^2\right\}$

est appelée *information* contenue dans l'échantillon (d'après R. Fisher). Désignons par $I_n(\vartheta)$ cette grandeur qui ne dépend que de la taille de l'échantillon et de la densité de probabilité $w_1(x; \vartheta)$ et supposons qu'elle ne soit pas nulle *). On obtient alors à partir de (2.39) la limite inférieure cherchée de la variance des estimations **)

$$M_2\{\hat{\vartheta}_n\} \geqslant \frac{[1 + b_n'(\vartheta)]^2}{I_n(\vartheta)}. \qquad (2.40)$$

Remarquons que le second membre de l'inégalité (2.40) est également la limite inférieure des écarts quadratiques moyens des estimations par rapport au paramètre estimé. Comme le minimum de la grandeur $m_1\{[\hat{\vartheta}_n - \vartheta]^2\}$ correspond à $m_1\{\hat{\vartheta}_n\} = \vartheta$ (cf. problème 2.6, tome I) on a

$$m_1\{[\hat{\vartheta}_n - \vartheta]^2\} \geqslant M_2\{\hat{\vartheta}_n\} \geqslant \frac{[1 + b_n'(\vartheta)]^2}{I_n(\vartheta)}. \qquad (2.41)$$

*) L'information est égale à zéro, par exemple, si la fonction de vraisemblance ne dépend pas de ϑ.

**) La formule (2.40) est appelée inégalité de Rao-Cramer et parfois inégalité de Fréchet. Ces trois auteurs ont obtenu ce résultat indépendamment.

Pour les estimations à biais constant ($b'(\vartheta) \equiv 0$) et, en particulier, pour les estimations non biaisées ($b(\vartheta) \equiv 0$) on a

$$M_2\{\hat{\vartheta}_n\} \geqslant \frac{1}{I_n(\vartheta)}. \tag{2.42}$$

Dans ce cas la limite inférieure de la variance des estimations est une grandeur inversement proportionnelle à l'information contenue dans l'échantillon. Les inégalités (2.40) à (2.42) deviennent des égalités pour les estimations efficaces.

Pour des échantillons de taille limitée, les cas où dans (2.42) il y a égalité sont extrêmement rares. Généralement la limite inférieure de la variance des estimations absolument correctes est plus grande que la grandeur inverse de l'information de Fisher. C'est pourquoi l'efficacité est la mesure non pas de la qualité de l'estimation, mais de la qualité de la limite inférieure dans l'inégalité de Rao-Cramer. Si la variance de l'estimation atteint exactement la limite inférieure, différente de $I_n^{-1}(\vartheta)$, l'écart par rapport à l'unité de l'efficacité ne peut servir d'indice significatif de l'insuffisance de l'estimation (pour plus de détail cf. [2], chap. VIII, ainsi que [11]).

Il est facile, en utilisant (2.10), d'exprimer l'information à l'aide de la fonction densité de probabilité $w_1(x;\ \vartheta)$. Comme

$$\ln W_n(x_1, \ldots, x_n \mid \vartheta) = \sum_{i=1}^{n} \ln w_1(x_i;\ \vartheta)$$

et

$$\frac{\partial}{\partial\vartheta} \ln W_n(x_1, \ldots, x_n \mid \vartheta) = \sum_{i=1}^{n} \frac{1}{w_1(x_i;\ \vartheta)} \frac{\partial}{\partial\vartheta} w_1(x_i;\ \vartheta),$$

compte tenu de (2.37) on a

$$I_n(\vartheta) = M_2\left\{\frac{\partial}{\partial\vartheta} \ln W_n\right\} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\vartheta} \ln W_n\right)^2 \times$$

$$\times W_n\, dx_1 \ldots dx_n = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \left[\sum_{i=1}^{n} \frac{1}{w_1(x_i;\ \vartheta)} \times\right.$$

$$\left.\times \frac{\partial w_1(x_i;\ \vartheta)}{\partial\vartheta}\right]^2 \prod_{i=1}^{n} w_1(x_i;\ \vartheta)\, dx_1 \ldots dx_n =$$

$$= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial\vartheta} \ln w_1(x_i;\ \vartheta)\right]^2 w_1(x_i;\ \vartheta)\, dx_i +$$

$$+ \sum_{i,\ j(i \neq j)}^{n} \int_{-\infty}^{\infty} \frac{\partial w_1(x_i;\ \vartheta)}{\partial\vartheta} dx_i \int_{-\infty}^{\infty} \frac{\partial w_1(x_j;\ \vartheta)}{\partial\vartheta} dx_j.$$

Comme les termes de la première somme sont égaux et

$$\int_{-\infty}^{\infty} \frac{\partial w_1(x_k;\ \vartheta)}{\partial \vartheta}\, dx_k = \frac{\partial}{\partial \vartheta} \int_{-\infty}^{\infty} w_1(x_k;\ \vartheta)\, dx_k = 0,$$

il vient

$$I_n(\vartheta) = n \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \vartheta} \ln w_1(x;\ \vartheta)\right]^2 w_1(x;\ \vartheta)\, dx =$$

$$= nm_1 \left\{\left[\frac{\partial}{\partial \vartheta} \ln w_1(x;\ \vartheta)\right]^2\right\} \qquad (2.43)$$

ou

$$I_n(\vartheta) = nI_1(\vartheta). \qquad (2.43')$$

Ainsi, l'information est proportionnelle à la taille de l'échantillon, le facteur de proportionnalité dépendant de la distribution initiale.

En substituant (2.43) dans (2.41) on obtient

$$m_1\{[\hat{\vartheta}_n - \vartheta]^2\} \geqslant M_2\{\hat{\vartheta}_n\} \geqslant$$

$$\geqslant \frac{1}{n} \frac{[1 + b_n'(\vartheta)]^2}{\int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \vartheta} \ln w_1(x;\ \vartheta)\right]^2 w_1(x;\ \vartheta)\, dx}. \qquad (2.44)$$

La grandeur $m_1\{[\hat{\vartheta}_n - \vartheta]^2\}^{1/2}$ sert parfois de mesure de la *précision* de l'estimation. Le second membre de l'inégalité (2.44) donne la *précision potentielle*.

Il faut remarquer que pour une distribution discrète l'expression (2.43) devient

$$I_n(\vartheta) = n \sum_i \left[\frac{\partial \ln p(x_i;\ \vartheta)}{\partial \vartheta}\right]^2 p(x_i;\ \vartheta). \qquad (2.45)$$

Supposons maintenant que l'estimation $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$ soit suffisante. On a alors en vertu de (2.28)

$$\frac{\partial}{\partial \vartheta} \ln W_n(x_1, \ldots, x_n \,|\, \vartheta) = \frac{\partial}{\partial \vartheta} \ln f(\hat{\vartheta}_n,\ \vartheta).$$

Supposons de plus que l'on ait

$$f(\hat{\vartheta}_n,\ \vartheta) = \exp\{k(\vartheta)\,[\hat{\vartheta}_n + c(\vartheta)]\} \qquad (2.46)$$

et, par conséquent,

$$\frac{\partial}{\partial \vartheta} \ln f(\hat{\vartheta}_n,\ \vartheta) = k'(\vartheta)\,\hat{\vartheta}_n + \frac{d}{d\vartheta}[k(\vartheta)\,c(\vartheta)].$$

On a alors en vertu de (2.37) et (2.46)

$$m_1\left\{\frac{\partial}{\partial \vartheta} \ln f(\hat{\vartheta}_n,\ \vartheta)\right\} = k'(\vartheta)\, m_1\{\hat{\vartheta}_n\} + \frac{d}{d\vartheta}[k(\vartheta)\,c(\vartheta)] = 0$$

et

$$\frac{\partial}{\partial \vartheta} \ln W_n = \frac{\partial}{\partial \vartheta} \ln f(\hat{\vartheta}_n,\ \vartheta) = k'(\vartheta)\,[\hat{\vartheta}_n - m_1\{\hat{\vartheta}_n\}]. \qquad (2.46')$$

En portant (2.46′) dans (2.38) on obtient

$$k'(\vartheta)\, m_1\{[\hat{\vartheta}_n - m_1\{\hat{\vartheta}_n\}]^2\} = k'(\vartheta)\, M_2\{\hat{\vartheta}_n\} = 1 + b_n'(\vartheta). \quad (2.47)$$

D'un autre côté [cf. (2.46′)]

$$I_n(\vartheta) = m_1\left\{\left(\frac{\partial}{\partial\vartheta}\ln W_n\right)^2\right\} = [k'(\vartheta)]^2 M_2\{\hat{\vartheta}_n\}. \quad (2.47')$$

A partir de (2.47) et (2.47′) on trouve

$$k'(\vartheta) = \frac{I_n(\vartheta)}{1+b_n'(\vartheta)}, \quad (2.47'')$$

et, par conséquent,

$$M_2\{\hat{\vartheta}_n\} = \frac{[1+b_n'(\vartheta)]^2}{I_n(\vartheta)}.$$

On a ainsi démontré que parmi toutes les estimations de biais donné $b_n(\vartheta)$ l'*estimation suffisante d'un certain type* [satisfaisant aux conditions (2.46) et (2.47″)] a une variance minimale [cf. (2.41)].

Ainsi, toute estimation efficace appartient à la classe des statistiques suffisantes, tandis que seules certaines des estimations suffisantes sont efficaces.

Dans ce paragraphe nous n'avons envisagé que des estimations conditionnelles. Cependant la notion d'efficacité peut être également étendue aux estimations inconditionnelles, ceci en prenant la moyenne de (2.30) sur ϑ.

L'inégalité de Rao-Cramer (2.40) s'écrira alors comme suit

$$\int_{-\infty}^{\infty} w_1(\vartheta)\, m_1\{[\hat{\vartheta}_n - \vartheta]^2\}\, d\vartheta \geqslant \frac{\int_{-\infty}^{\infty} w_1(\vartheta)\,[1+b_n'(\vartheta)]^2\, d\vartheta}{\int_{-\infty}^{\infty} w_1(\vartheta)\, I_n(\vartheta)\, d\vartheta}.$$

2.2.6. Estimations d'intervalle. On entend par estimation d'intervalle du paramètre ϑ l'intervalle dont les limites $\hat{\vartheta}_n^{(\inf)}$ et $\hat{\vartheta}_n^{(\sup)}$ sont des fonctions des valeurs échantillonnées $x_1, \ldots, x_n$ et qui contient le paramètre estimé avec une certaine probabilité. Sous la forme analytique ceci peut s'écrire comme suit:

$$P\{\hat{\vartheta}_n^{(\inf)} < \vartheta < \hat{\vartheta}_n^{(\sup)}\} = \gamma. \quad (2.48)$$

La probabilité γ est appelée *coefficient de confiance* et les estimations $\hat{\vartheta}_n^{(\inf)}$ et $\hat{\vartheta}_n^{(\sup)}$ sont respectivement les *limites* inférieure et supérieure *de confiance*. L'intervalle $(\hat{\vartheta}_n^{(\inf)}, \hat{\vartheta}_n^{(\sup)})$ est appelé *intervalle de confiance.*

Parfois l'intervalle de confiance est défini comme suit:

$$\hat{\vartheta}_n^{(\inf)} = \hat{\vartheta}_n - \varepsilon_1\vartheta, \quad (2.49)$$

$$\hat{\vartheta}_n^{(\sup)} = \hat{\vartheta}_n + \varepsilon_2\vartheta, \quad (2.49')$$

où $\hat{\vartheta}_n$ est l'estimation ponctuelle du paramètre ϑ; ε_1 et ε_2 sont des nombres positifs. Pour une valeur donnée de γ, celles de ε_1 et ε_2 peuvent être calculées d'une infinité de manières. Si $W_{\hat{\vartheta}_n}(z)$ est la densité de probabilité d'une estimation ponctuelle, on a en vertu de (2.48), (2.49) et (2.49')

$$P\,\{\vartheta\,(1-\varepsilon_2) < \hat{\vartheta}_n < \vartheta\,(1+\varepsilon_1)\} = \gamma, \tag{2.50}$$

on en tire deux relations donnant ε_1 et ε_2:

$$\int_{-\infty}^{\vartheta(1-\varepsilon_2)} W_{\hat{\vartheta}_n}(z)\,dz = \gamma_2, \tag{2.51}$$

$$\int_{\vartheta(1+\varepsilon_1)}^{\infty} W_{\hat{\vartheta}_n}(z)\,dz = \gamma_1, \tag{2.51'}$$

où γ_1, γ_2 sont des nombres positifs quelconques inférieurs à l'unité, avec de plus $\gamma_1 + \gamma_2 = 1 - \gamma$.

Pour $\varepsilon_1 = \varepsilon_2 = \varepsilon$ la formule (2.50) donne la relation entre le coefficient de confiance γ, la longueur relative 2ε de l'intervalle de confiance et la taille de l'échantillon n:

$$P\{\vartheta\,(1-\varepsilon) < \hat{\vartheta}_n < \vartheta\,(1+\varepsilon)\} = \int_{\vartheta(1-\varepsilon)}^{\vartheta(1+\varepsilon)} W_{\hat{\vartheta}_n}(z)\,dz = \gamma. \tag{2.52}$$

Si l'on se donne la longueur de l'intervalle de confiance, pour les estimations consistantes et sans biais le coefficient de confiance croît avec la taille de l'échantillon, tendant vers l'unité. Pour une taille donnée de l'échantillon, le coefficient de confiance sera d'autant plus grand que sera large l'intervalle de confiance. En d'autres termes, pour une taille donnée de l'échantillon, il est impossible d'élever le coefficient de confiance sans augmenter la longueur de l'intervalle de confiance ni de restreindre cet intervalle sans réduire le coefficient de confiance.

Trois types de problèmes utilisant les estimations d'intervalle d'un paramètre peuvent être formulés. Pour un échantillon de taille donnée n on prend l'estimation ponctuelle $\hat{\vartheta}_n$, on calcule sa distribution $W_{\hat{\vartheta}_n}(z)$ et pour une valeur donnée de ε on trouve de (2.52) le coefficient de confiance γ. Dans les mêmes conditions on peut, pour un γ donné, trouver la longueur relative 2ε de l'intervalle de confiance. Enfin, on peut se donner le coefficient de confiance γ et la longueur relative 2ε de l'intervalle de confiance. On trouve alors à partir de (2.52), par approximations successives, la taille de l'échantillon laquelle permet d'avoir simultanément γ et ε donnés.

2.3. MÉTHODES D'OBTENTION DES ESTIMATIONS POUR DES CRITÈRES DONNÉS

2.3.1. Estimation du maximum de vraisemblance. La fonction de vraisemblance d'un échantillon de taille n pour des valeurs échantillonnées $x_1, \ldots, x_n$ est une fonction du paramètre inconnu ϑ [cf. (2.10)]

$$W_n(x_1, \ldots, x_n \mid \vartheta) = \prod_{k=1}^{n} w_1(x_k\,;\,\vartheta).$$

Pour souligner ceci, introduisons une nouvelle désignation pour la fonction de vraisemblance, soit:

$$L_x(\vartheta) = W_n(x_1, \ldots, x_n \mid \vartheta). \tag{2.53}$$

Choisissons pour l'estimation conditionnelle du paramètre inconnu la valeur de la variable ϑ pour laquelle, pour des $x_1, \ldots, x_n$ donnés, la fonction de vraisemblance $L_x(\vartheta)$ atteint un maximum. L'estimation satisfaisant au critère de qualité formulé est dite *estimation du maximum de vraisemblance.*

La condition d'extrémum de la fonction de vraisemblance s'écrit comme suit:

$$\frac{\partial L_x(\vartheta)}{\partial \vartheta} = 0. \tag{2.54}$$

Cependant, dans la pratique, on n'utilise pas l'équation (2.54). Un artifice mathématique simple permet de simplifier cette équation. En effet, le logarithme étant une fonction monotone, les extréma des fonctions $L_x(\vartheta)$ et $\ln L_x(\vartheta)$ correspondent aux mêmes valeurs de l'argument ϑ. L'équation du maximum de vraisemblance s'écrit alors

$$\frac{\partial \ln L_x(\vartheta)}{\partial \vartheta} = 0 \tag{2.55}$$

ou, compte tenu de (2.10),

$$\sum_{k=1}^{n} \frac{\partial \ln w_1(x_k\,;\,\vartheta)}{\partial \vartheta} = 0. \tag{2.56}$$

Dans le cas d'une distribution discrète, compte tenu de (2.10′), on obtient

$$\sum_{k=1}^{n} \frac{\partial \ln p(x_k\,;\,\vartheta)}{\partial \vartheta} = 0. \tag{2.56′}$$

L'équation du maximum de vraisemblance est en général une équation non linéaire, algébrique ou transcendante, pouvant avoir plusieurs solutions correspondant aux maxima et aux minima relatifs de la fonction de vraisemblance. Chacune des solutions $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$, correspondant à un maximum de la fonction de vraisemblance, est une estimation du maximum de vraisemblance du paramètre inconnu. Le problème se réduit dans ce

cas à la recherche de la solution correspondant au maximum absolu de la fonction de vraisemblance.

Si pour une taille donnée n de l'échantillon, l'estimation conditionnelle efficace non biaisée $\hat{\vartheta}_{n\,\mathrm{eff}}$ du paramètre ϑ existe, l'équation de vraisemblance (2.55) peut en vertu de (2.46′) s'écrire comme suit :

$$\frac{\partial \ln L_x(\vartheta)}{\partial \vartheta} = k'(\vartheta)\,[\hat{\vartheta}_{n\,\mathrm{eff}} - \vartheta] = 0,$$

où conformément à (2.47″) on a

$$k'(\vartheta) = I_n(\vartheta) > 0.$$

Comme $\frac{\partial \ln L_x(\vartheta)}{\partial \vartheta} > 0$ pour $\vartheta < \hat{\vartheta}_{n\,\mathrm{eff}}$ et $\frac{\partial \ln L_x(\vartheta)}{\partial \vartheta} < 0$ pour $\vartheta > \hat{\vartheta}_{n\,\mathrm{eff}}$, l'estimation du maximum de vraisemblance est égale à

$$\hat{\vartheta}_n = \hat{\vartheta}_{n\,\mathrm{eff}}.$$

On en tire une conclusion fort importante.

Si l'estimation efficace non biaisée $\hat{\vartheta}_{n\,\mathrm{eff}}$ existe, l'équation du maximum de vraisemblance a une solution *unique* égale à $\hat{\vartheta}_{n\,\mathrm{eff}}$.

Si l'estimation suffisante $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$ existe, en vertu de (2.28) et (2.55) l'équation de vraisemblance peut s'écrire comme suit :

$$\frac{\partial \ln L_x(\vartheta)}{\partial \vartheta} = \frac{\partial \ln f[g(x_1, \ldots, x_n), \vartheta]}{\partial \vartheta} = 0.$$

Par conséquent, s'il existe une estimation suffisante, chaque solution de l'équation de vraisemblance est une fonction de cette estimation suffisante.

On peut montrer (cf. [5], § 33) que, pour des restrictions pas trop fortes relatives à la fonction $w_1(x; \vartheta)$, l'estimation du maximum de vraisemblance est *consistante*. Cette estimation est *asymptotiquement correcte et asymptotiquement efficace*. De plus, la distribution de l'estimation du maximum de vraisemblance est asymptotiquement normale de moyenne ϑ et de variance égale à l'inverse de la valeur de l'information [cf. (2.48)], c'est-à-dire pour $n \to \infty$

$$W_{\hat{\vartheta}_n}(z) \sim \sqrt{\frac{I_n(\vartheta)}{2\pi}} \exp\left[-\frac{I_n(\vartheta)}{2}(z - \vartheta)^2\right], \qquad (2.57)$$

de plus, $I_n(\vartheta) \to \infty$ pour $n \to \infty$.

2.3.2. Solution approchée de l'équation de vraisemblance. Des méthodes simples ne permettent pas toujours de trouver la solution exacte de l'équation de vraisemblance (2.55). Dans le cas général, on peut recommander la méthode des approximations successives *).

*) C'est une variante de la méthode itérative de Newton. Pour plus de détails voir, par exemple, E. Boot. *Méthodes numériques* (en russe). Ed. « Physmathguiz », Moscou, 1959, page 187.

En développant le logarithme de la fonction de vraisemblance en série de Taylor au voisinage du point ϑ_0 et en ne gardant que les trois premiers termes on a

$$\ln L_x(\vartheta) \approx \ln L_x(\vartheta_0) + (\vartheta - \vartheta_0)\left[\frac{\partial}{\partial\vartheta}\ln L_x(\vartheta)\right]_{\vartheta=\vartheta_0} + \\ + \frac{1}{2}(\vartheta-\vartheta_0)^2\left[\frac{\partial^2}{\partial\vartheta^2}\ln L_x(\vartheta)\right]_{\vartheta=\vartheta_0}. \quad (2.58)$$

A partir de (2.58) on trouve directement

$$\frac{\partial \ln L_x(\vartheta)}{\partial\vartheta} \approx \left[\frac{\partial}{\partial\vartheta}\ln L_x(\vartheta)\right]_{\vartheta=\vartheta_0} + (\vartheta-\vartheta_0)\left[\frac{\partial^2}{\partial\vartheta^2}\ln L_x(\vartheta)\right]_{\vartheta=\vartheta_0}. \quad (2.59)$$

Prenons pour première approximation de la solution de l'équation du maximum de vraisemblance *) une estimation grossière quelconque $\hat{\vartheta}_1$ du paramètre inconnu ϑ. Ce peut être la moyenne échantillonnée, ou la médiane échantillonnée, ou toute autre estimation. Utilisons maintenant (2.59) pour le calcul de la seconde approximation $\hat{\vartheta}_2$ en tant que racine de l'équation

$$\left[\frac{\partial}{\partial\vartheta}\ln L_x(\vartheta)\right]_{\vartheta=\hat{\vartheta}_1} + (\vartheta-\hat{\vartheta}_1)\left[\frac{\partial^2}{\partial\vartheta^2}\ln L_x(\vartheta)\right]_{\vartheta=\hat{\vartheta}_1} = 0. \quad (2.60)$$

En vertu de (2.60) on a

$$\hat{\vartheta}_2 = \hat{\vartheta}_1 - \frac{\left[\frac{\partial}{\partial\vartheta}\ln L_x(\vartheta)\right]_{\vartheta=\hat{\vartheta}_1}}{\left[\frac{\partial^2}{\partial\vartheta^2}\ln L_x(\vartheta)\right]_{\vartheta=\hat{\vartheta}_1}}. \quad (2.61)$$

En remplaçant $\frac{\partial^2}{\partial\vartheta^2}\ln L_x(\vartheta)$ par sa valeur moyenne (prise sur l'espace des échantillons) il vient

$$\frac{\partial^2}{\partial\vartheta^2}\ln L_x(\vartheta) \approx m_1\left\{\frac{\partial^2}{\partial\vartheta^2}\ln L_x(\vartheta)\right\} = \\ = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial^2}{\partial\vartheta^2}[\ln L_x(\vartheta)]\,L_x(\vartheta)\,dx_1\ldots dx_n = \\ = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\frac{\partial^2 L_x(\vartheta)}{\partial\vartheta^2} - \left[\frac{1}{L_x(\vartheta)}\frac{\partial L_x(\vartheta)}{\partial\vartheta}\right]^2 L_x(\vartheta)\right\}dx_1\ldots dx_n. \quad (2.62)$$

En vertu de (2.37) on obtient

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial^2 L_x(\vartheta)}{\partial\vartheta^2}\,dx_1\ldots dx_n = 0. \quad (2.63)$$

*) Nous supposons ici que la fonction de vraisemblance est unimodale ou que la première approximation se trouve au voisinage du maximum principal de cette fonction.

Et à partir de (2.62)

$$\frac{\partial^2}{\partial \vartheta^2} \ln L_x(\vartheta) \approx m_1 \left\{ \frac{\partial^2}{\partial \vartheta^2} \ln L_x(\vartheta) \right\} =$$
$$= -m_1 \left\{ \left[\frac{\partial \ln L_x(\vartheta)}{\partial \vartheta} \right]^2 \right\} = -I_n(\vartheta). \qquad (2.64)$$

En portant (2.64) dans (2.61) on trouve l'expression de la seconde approximation, soit :

$$\hat{\vartheta}_2 = \hat{\vartheta}_1 + \frac{1}{I_n(\hat{\vartheta}_1)} \left[\frac{\partial}{\partial \vartheta} \ln L_x(\vartheta) \right]_{\vartheta = \hat{\vartheta}_1}. \qquad (2.65)$$

On peut obtenir l'approximation suivante, en remplaçant ϑ_0 dans le second membre de (2.59) par $\hat{\vartheta}_2$ donnée par (2.65), en égalant à zéro ce second membre, en trouvant la racine de l'équation et en y remplaçant de plus la dérivée seconde négative par sa valeur moyenne, qui est l'information I_n $(\hat{\vartheta}_2)$. Une fois la m-ième approximation trouvée, l'approximation suivante est donnée par la formule

$$\hat{\vartheta}_{m+1} = \hat{\vartheta}_m + \frac{1}{I_n(\hat{\vartheta}_m)} \left[\frac{\partial}{\partial \vartheta} \ln L_x(\vartheta) \right]_{\vartheta = \hat{\vartheta}_m} \qquad (2.66)$$

qui est une généralisation évidente de (2.65).

2.3.3. Maximum de densité de probabilité a posteriori du paramètre estimé. Supposons maintenant que le paramètre inconnu ϑ soit une variable aléatoire dont la densité de probabilité est w_1 (ϑ). A l'aide de la formule de Bayes [cf. (2.41), tome I] on peut trouver la densité de probabilité conditionnelle du paramètre pour un échantillon donné $x_1, \ldots, x_n$:

$$W_1(\vartheta \mid x_1, \ldots, x_n) = \frac{w_1(\vartheta) L_x(\vartheta)}{\int\limits_{-\infty}^{\infty} w_1(\vartheta) L_x(\vartheta) \, d\vartheta}. \qquad (2.67)$$

Choisissons en qualité d'estimation du paramètre la valeur de la variable ϑ pour laquelle, pour des $x_1, \ldots, x_n$ donnés, la densité de probabilité a posteriori du paramètre estimé est maximale. Compte tenu de la remarque faite à propos de l'équation du maximum de vraisemblance (2.55), on peut écrire sous la forme suivante l'équation de l'estimation correspondant au maximum de la densité de probabilité a posteriori :

$$\frac{\partial}{\partial \vartheta} \ln [w_1(\vartheta) L_x(\vartheta)] = 0 \qquad (2.68)$$

ou

$$\frac{\partial}{\partial \vartheta} \ln w_1(\vartheta) + \sum_{k=1}^{n} \frac{\partial}{\partial \vartheta} \ln w_1(x_k \mid \vartheta) = 0. \qquad (2.69)$$

Les estimations envisagées ont des propriétés voisines de celles des estimations conditionnelles du maximum de vraisemblance (cf. [9], § 21.1.2). Si $\hat{\Theta}_{n\,\text{eff}}$ est l'estimation inconditionnelle efficace et non biaisée du paramètre ϑ, l'équation (2.68) a une solution unique égale à $\hat{\Theta}_{n\,\text{eff}}$. Pour des restrictions assez faibles, l'estimation correspondant au maximum de densité de probabilité a posteriori du paramètre estimé est consistante et asymptotiquement efficace. Pour $n \to \infty$ la densité de probabilité de cette estimation peut être représentée par la relation asymptotique suivante:

$$W_{\hat{\vartheta}_n}(z) \sim \sqrt{\frac{m_1\{I_n(\vartheta)\}}{2\pi}} \exp\left\{-\frac{m_1\{I_n(\vartheta)\}}{2}[z-m_1\{\vartheta\}]^2\right\}. \quad (2.70)$$

Si la distribution a priori du paramètre ϑ est exactement ou à peu près uniforme sur un certain intervalle fini, on a en vertu de (2.67)

$$W_1(\vartheta \mid x_1, \ldots, x_n) = kL_x(\vartheta),$$

où k est une constante ne dépendant pas de ϑ.

Dans ce cas les estimations du maximum de vraisemblance seront en même temps des estimations correspondant au maximum de la densité de probabilité a posteriori du paramètre estimé ϑ.

2.3.4. Estimations de Bayes. Les estimations ponctuelles obtenues pour les critères du maximum de vraisemblance ou de densité de probabilité a posteriori répondent aux exigences d'un type spécial et peuvent évaluer mal les pertes dues aux estimations peu heureuses. En effet, à toute estimation $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$ calculée à partir de l'échantillon observé $x_1, \ldots, x_n$ correspond une *erreur*, mesurée par la différence $\vartheta - \hat{\vartheta}_n$. Cette différence variera d'un échantillon à l'autre, parfois elle peut être insignifiante et parfois très importante. Pour en tenir compte on associe à chaque erreur un certain poids et l'on introduit une *fonction de pertes* non négative $\Pi(\hat{\vartheta}_n, \vartheta)$. Dans le cas général cette fonction peut dépendre non seulement de la différence $\vartheta - \hat{\vartheta}_n$, mais également de chacune des grandeurs figurant dans son argument.

Le choix de la fonction de pertes est en certain sens arbitraire et dépend du cas envisagé. Le plus souvent, dans les applications, on rencontre les fonctions suivantes (voir fig. 2.3):

la fonction de pertes simple *)

$$\Pi(\hat{\vartheta}_n, \vartheta) = c - \delta(\hat{\vartheta}_n - \vartheta), \quad c > 0; \quad (2.71)$$

la fonction de pertes égale au module de l'erreur

$$\Pi(\hat{\vartheta}_n, \vartheta) = |\hat{\vartheta}_n - \vartheta|; \quad (2.72)$$

*) La fonction (2.71) n'est pas négative à l'exception du point $\hat{\vartheta}_n = \vartheta$.

la fonction de pertes quadratique

$$\Pi(\hat{\vartheta}_n, \vartheta) = (\hat{\vartheta}_n - \vartheta)^2; \qquad (2.73)$$

la fonction de pertes rectangulaire

$$\Pi(\hat{\vartheta}_n, \vartheta) = \begin{cases} 0, & |\vartheta - \hat{\vartheta}_n| < \varepsilon, \\ 1, & |\vartheta - \hat{\vartheta}_n| \geqslant \varepsilon; \end{cases} \qquad (2.74)$$

la fonction de pertes exponentielle

$$\Pi(\hat{\vartheta}_n, \vartheta) = 1 - \exp[-k(\hat{\vartheta}_n - \vartheta)^2]. \qquad (2.75)$$

Toutes les fonctions de pertes mentionnées sont des fonctions symétriques de l'erreur $\hat{\vartheta}_n - \vartheta$.

On peut utiliser comme mesure de la qualité d'une estimation conditionnelle la moyenne des pertes prise sur tous les échantillons

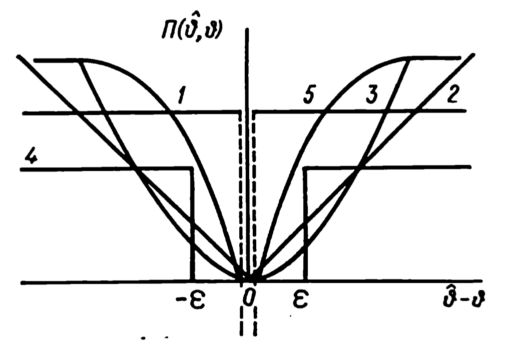


Fig. 2.3. Fonctions de pertes:
1 — simple; 2 — égale au module de l'erreur; 3 — quadratique; 4 — rectangulaire; 5 — exponentielle

possibles, compte tenu de la fréquence d'apparition des différents échantillons, c'est-à-dire de la fonction de vraisemblance $L_x(\vartheta)$. Cette moyenne égale à

$$r(\vartheta) = m_1\{\Pi(\hat{\vartheta}_n, \vartheta)\} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \Pi(\hat{\vartheta}_n, \vartheta) L_x(\vartheta)\, dx_1 \dots dx_n, \qquad (2.76)$$

est appelée *fonction conditionnelle de risque*. Demandons maintenant que l'estimation conditionnelle $\hat{\vartheta}_n$ minimise la fonctionnelle (2.76). L'estimation, optimale au point de vue du critère du minimum de la fonction de risque, est appelée *estimation conditionnelle de Bayes*.

Si le paramètre estimé ϑ est aléatoire et $w_1(\vartheta)$ est sa densité de probabilité, en prenant la moyenne de la fonction conditionnelle

de risque sur toutes les valeurs éventuelles de ϑ, on obtient le *risque moyen*

$$R = \int_{-\infty}^{\infty} w_1(\vartheta)\, r(\vartheta)\, d\vartheta. \tag{2.77}$$

En portant dans (2.77) l'expression de $r(\vartheta)$ à partir de (2.76) et compte tenu de (2.67) on peut écrire le risque moyen sous la forme suivante:

$$R = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} J(\hat{\vartheta}_n \mid x_1, \ldots, x_n)\, W_n(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \\ = m_1\{J(\hat{\vartheta}_n \mid x_1, \ldots, x_n)\}, \tag{2.78}$$

$$J(\hat{\vartheta}_n \mid x_1, \ldots, x_n) = \int_{-\infty}^{\infty} \Pi(\hat{\vartheta}_n, \vartheta)\, W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta; \tag{2.79}$$

$$W_n(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} w_1(\vartheta)\, L_x(\vartheta)\, d\vartheta. \tag{2.79'}$$

Les estimations obtenues d'après le critère du minimum de risque moyen sont appelées *estimations inconditionnelles de Bayes* (ou simplement estimations de Bayes).

L'estimation de Bayes pour une fonction de pertes et une distribution a priori du paramètre ϑ données peut être trouvée à partir de la condition du minimum de la fonctionnelle (2.78) dépendant de la forme de la fonction $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$. Comme la fonctionnelle J dépend de g et ne dépend pas des dérivées partielles $\frac{\partial g}{\partial x_i}$, l'équation d'Euler bien connue dans le calcul des variations, utilisée pour la détermination d'une fonction extrémale s'écrira comme suit:

$$\frac{\partial J}{\partial g} = 0 \tag{2.80}$$

à condition toutefois que la dérivée dans (2.80) existe.

2.3.5. Fonction de pertes simple. Nous allons écrire l'expression de la fonction conditionnelle de risque pour la fonction de pertes simple. En portant (2.71) dans (2.76) et en utilisant la propriété de filtrage de la fonction delta dans l'intégrale multiple [cf. (3.117), tome I], on obtient

$$r(\vartheta) = c - L_{\hat{\vartheta}_n}(\vartheta), \tag{2.81}$$

où $L_{\hat{\vartheta}_n}(\vartheta)$ est la densité de probabilité de l'estimation $\hat{\vartheta}_n$ pour ϑ donné. Pour que $\hat{\vartheta}_n$ soit une estimation conditionnelle de Bayes, elle doit minimiser la fonction de risque $r(\vartheta)$, ce qui équivaut à

exiger que

$$L_{\hat{\vartheta}_n^*}(\vartheta) \geqslant L_{\hat{\vartheta}_n}(\vartheta). \qquad (2.82)$$

Cette expression signifie que $\hat{\vartheta}_n^*$ est une estimation conditionnelle du maximum de vraisemblance.

Ainsi, les estimations conditionnelles du maximum de vraisemblance sont une forme particulière des *estimations conditionnelles de Bayes pour une fonction de pertes simple.*

En portant (2.81) dans (2.77), on trouve également l'expression du risque moyen pour la fonction de pertes simple, soit :

$$R = c - \int_{-\infty}^{\infty} w_1(\vartheta) L_{\hat{\vartheta}_n}(\vartheta)\, d\vartheta, \qquad (2.83)$$

en vertu de laquelle l'estimation $\hat{\vartheta}_n^*$ est une estimation inconditionnelle de Bayes si

$$w_1(\vartheta) L_{\hat{\vartheta}_n^*}(\vartheta) \geqslant w_1(\vartheta) L_{\hat{\vartheta}_n}(\vartheta). \qquad (2.84)$$

Ainsi, l'estimation correspondant au maximum de densité de probabilité a posteriori [cf. (2.67)] est une forme particulière des *estimations inconditionnelles de Bayes pour une fonction de pertes simple.*

2.3.6. Fonction de pertes quadratique. La fonction de risque conditionnelle est égale à

$$r(\vartheta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\hat{\vartheta}_n - \vartheta)^2 L_x(\vartheta)\, dx_1 \ldots dx_n, \qquad (2.85)$$

c'est-à-dire qu'elle coïncide avec l'écart quadratique moyen de l'estimation par rapport au paramètre estimé (ou avec la variance de l'estimation, si celle-ci est absolument correcte). C'est pourquoi la condition du minimum de la fonction de risque $r(\vartheta)$ se réduit à la minimisation de $m_1\{(\hat{\vartheta}_n - \vartheta)^2\}$, c'est-à-dire à la recherche de l'estimation efficace (cf. § 2.2.5). Ainsi, *l'estimation efficace (conditionnelle) est une estimation de Bayes conditionnelle pour une fonction de pertes quadratique.* En portant (2.73) dans (2.79) on trouve la fonctionnelle $J(\hat{\vartheta}_n \mid x_1, \ldots, x_n)$ dans l'expression du risque moyen (2.78)

$$J(\hat{\vartheta}_n \mid x_1, \ldots, x_n) = \int_{-\infty}^{\infty} (\hat{\vartheta}_n - \vartheta)^2 W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta. \qquad (2.86)$$

En portant alors (2.86) dans (2.80) et en résolvant l'équation par rapport à la fonction inconnue $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$, on trouve

l'expression cherchée de l'estimation de Bayes, soit *):

$$\hat{\vartheta}_n^* = g^*(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} \vartheta W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta \qquad (2.87)$$

ou

$$\hat{\vartheta}_n^* = m_1\{\vartheta \mid x_1, \ldots, x_n\} = \frac{\int_{-\infty}^{\infty} \vartheta w_1(\vartheta)\, L_x(\vartheta)\, d\vartheta}{\int_{-\infty}^{\infty} w_1(\vartheta)\, L_x(\vartheta)\, d\vartheta}. \qquad (2.87')$$

En vertu de (2.87) l'estimation de Bayes pour une fonction de pertes quadratique coïncide avec la *moyenne conditionnelle* du paramètre estimé ϑ pour des $x_1, \ldots, x_n$ donnés.

En portant (2.87) dans (2.86) on trouve (pour une estimation absolument correcte)

$$J^* = J(\hat{\vartheta}_n^* \mid x_1, \ldots, x_n) = M_2\{\vartheta \mid x_1, \ldots, x_n\},$$

et, par conséquent, le risque minimal (bayesien) pour une fonction de pertes quadratique est égal à

$$R^* = m_1\{M_2\{\vartheta \mid x_1, \ldots, x_n\}\} = M_2\{\vartheta\}. \qquad (2.88)$$

A la différence de la fonction de pertes simple, pour laquelle l'estimation de Bayes est déterminée par les propriétés locales de la densité de probabilité a posteriori du paramètre estimé ϑ au voisinage de son maximum, l'estimation de Bayes pour une fonction de pertes quadratique dépend de la variation de la densité a posteriori mentionnée, dans toute la gamme de variation du paramètre ϑ. Il faut cependant remarquer que pour une densité de probabilité a posteriori $W_1(\vartheta \mid x_1, \ldots, x_n)$ unimodale et symétrique par rapport au mode, l'estimation de Bayes pour une fonction de pertes quadratique coïncide en vertu de (2.87) avec l'estimation suivant le critère du maximum de densité a posteriori, qui est l'estimation de Bayes pour une fonction de pertes simple.

2.3.7. Fonction de pertes égale au module de l'erreur. Pour la fonction de pertes (2.72) on trouve que la fonction $J(\hat{\vartheta}_n \mid x_1, \ldots, x_n)$ dans l'expression du risque moyen (2.78) est

$$J(\hat{\vartheta}_n \mid x_1, \ldots, x_n) = \int_{-\infty}^{\infty} |\hat{\vartheta}_n - \vartheta|\, W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta =$$

$$= \int_{-\infty}^{\hat{\vartheta}_n} (\hat{\vartheta}_n - \vartheta)\, W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta -$$

$$- \int_{\hat{\vartheta}_n}^{\infty} (\hat{\vartheta}_n - \vartheta)\, W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta. \qquad (2.89)$$

*) Il est facile de voir que (2.87) correspond au *minimum* du risque moyen R.

En portant (2.89) dans (2.80) on trouve

$$\frac{\partial J}{\partial \hat{\vartheta}_n} = \int_{-\infty}^{\hat{\vartheta}_n} W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta - \int_{\hat{\vartheta}_n}^{\infty} W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta = 0$$

ou

$$\int_{-\infty}^{\hat{\vartheta}_n} W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta = \int_{\hat{\vartheta}_n}^{\infty} W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta. \qquad (2.89')$$

L'égalité (2.89′) montre que l'estimation de Bayes, pour une fonction de pertes égale au module de l'erreur, coïncide avec la *médiane conditionnelle* du paramètre estimé ϑ pour des $x_1, \ldots, x_n$ donnés.

Si la densité de probabilité a posteriori du paramètre estimé est une fonction unimodale et symétrique par rapport au mode, la médiane et la moyenne de cette fonction coïncident et sont égales à son mode. Dans ce cas les estimations de Bayes, pour une fonction de pertes égale au module de l'erreur et une fonction de pertes quadratique, sont les mêmes et coïncident avec l'estimation du maximum de probabilité a posteriori.

2.3.8. Fonction de pertes rectangulaire. Pour la fonction de pertes (2.74) on trouve de (2.79) que la fonction $J(\hat{\vartheta}_n \mid x_1, \ldots, x_n)$ dans l'expression du risque moyen (2.78) est

$$J(\hat{\vartheta}_n \mid x_1, \ldots, x_n) = \int_{-\infty}^{\hat{\vartheta}_n - \varepsilon} W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta + \int_{\hat{\vartheta}_n + \varepsilon}^{\infty} W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta = 1 - \int_{\hat{\vartheta}_n - \varepsilon}^{\hat{\vartheta}_n + \varepsilon} W_1(\vartheta \mid x_1, \ldots, x_n)\, d\vartheta. \qquad (2.90)$$

En portant (2.90) dans (2.80) on obtient pour l'estimation de Bayes dans le cas d'une fonction de pertes rectangulaire l'équation transcendante suivante:

$$W_1(\hat{\vartheta}_n + \varepsilon \mid x_1, \ldots, x_n) = W_1(\hat{\vartheta}_n - \varepsilon \mid x_1, \ldots, x_n). \qquad (2.90')$$

Si la densité de probabilité a posteriori du paramètre estimé est une fonction unimodale et symétrique par rapport au mode, l'unique solution de (2.90′) est l'estimation $\hat{\vartheta}_n$ coïncidant avec le mode de cette fonction. Par conséquent, dans ce cas l'estimation de Bayes pour une fonction de pertes rectangulaire coïncide avec l'estimation correspondant au maximum de probabilité a posteriori (c'est-à-dire avec l'estimation de Bayes pour des fonctions de pertes simple et

quadratique). L'estimation conditionnelle de Bayes sera alors l'estimation du maximum de vraisemblance.

2.3.9. Fonction de pertes symétrique. Soit une fonction de pertes dérivable arbitraire, paire par rapport à l'erreur, satisfaisant donc à la condition suivante:

$$\Pi(\hat{\vartheta}_n - \vartheta) = \Pi(\vartheta - \hat{\vartheta}_n). \tag{2.91}$$

Supposons que la densité de probabilité conditionnelle du paramètre ϑ pour des $x_1, \ldots, x_n$ donnés soit unimodale et symétrique par rapport au mode. En vertu de cette hypothèse, la moyenne conditionnelle est le mode de la densité conditionnelle, c'est-à-dire que $W_1(\vartheta - m_1\{\vartheta \mid x_1, \ldots, x_n\} \mid x_1, \ldots, x_n)$ est une fonction paire de l'argument $\vartheta - m_1$. Ecrivons maintenant l'expression de la dérivée de la fonctionnelle (2.79)

$$\frac{\partial J}{\partial g} = \int_{-\infty}^{\infty} \frac{\partial}{\partial g} [\Pi(g-\vartheta)] W_1(\vartheta - m_1\{\vartheta \mid x_1, \ldots, x_n\} \mid x_1, \ldots, x_n)\, d\vartheta. \tag{2.92}$$

Comme $\Pi(g - \vartheta)$ est une fonction paire, sa dérivée est une fonction impaire de l'argument $g - \vartheta$. Le second facteur de la fonction sous l'intégrale (2.92) étant une fonction paire de $\vartheta - m_1\{\vartheta \mid x_1, \ldots, x_n\}$, la grandeur $\frac{\partial J}{\partial g}$ est identiquement nulle, si l'estimation $\hat{\vartheta}_n = g(x_1, \ldots, x_n)$ est égale à

$$\hat{\vartheta}_n = m_1\{\vartheta \mid x_1, \ldots, x_n\}. \tag{2.93}$$

En effet, lorsque l'égalité (2.93) se trouve vérifiée, la fonction sous l'intégrale devient une fonction impaire par rapport à la nouvelle variable d'intégration $\tau = \vartheta - m_1\{\vartheta \mid x_1, \ldots, x_n\}$.

Ainsi, l'estimation (2.93) est une solution de l'équation (2.80) et, par conséquent, une *estimation de Bayes.*

En comparant (2.93) et (2.87'), on voit que l'estimation de Bayes, pour une fonction de pertes quadratique (quelle que soit la densité de probabilité a posteriori du paramètre estimé), est également une estimation de Bayes pour une fonction de pertes symétrique, si la densité de probabilité a posteriori du paramètre estimé est unimodale et symétrique par rapport au mode.

2.3.10. Estimations du minimax. Lorsque la distribution a priori $w_1(\vartheta)$ n'est pas connue, on peut essayer d'obtenir l'estimation inconditionnelle du paramètre aléatoire ϑ à partir du critère du *minimax.* Soit une famille de courbes de fonctions conditionnelles de risque correspondant à différentes estimations $\hat{\vartheta}_n$ (fig. 2.4). On appelle estimation du minimax $\hat{\vartheta}_{n\,\mathrm{mm}}$ l'estimation pour laquelle le supremum de la fonction conditionnelle de risque $r(\vartheta)$ n'est pas supérieur à ceux de la même fonction (par rapport à la variable ϑ)

pour toute autre estimation. Tout comme le critère du minimax en cas d'alternatives, l'estimation du minimax nous garantit que les pertes en moyenne (sur l'ensemble des échantillons de taille donnée) ne seront pas supérieures à une certaine valeur r_{min}. Souvent l'estimation du minimax peut s'avérer trop prudente (voir fig. 1.1).

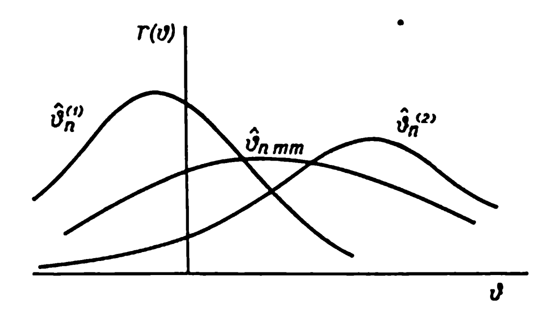


Fig. 2.4. Famille de courbes des fonctions de risque conditionnelles pour différentes estimations

Comme l'a montré A. Wald [14], l'estimation du minimax est une estimation de Bayes pour la distribution a priori la moins favorable du paramètre estimé, pour laquelle le risque moyen (pour une fonction de pertes donnée) est maximal. Il a montré également que l'estimation de Bayes pour laquelle la fonction conditionnelle de risque devient une constante (c'est-à-dire ne dépend pas de ϑ) est une estimation du minimax. Cependant, si cette fonction devient une constante pour une estimation non bayesienne, ceci ne signifie pas que l'estimation est du minimax. D'un autre côté, si la condition $r(\vartheta) = \text{const}$ n'est pas réalisable pour tous les ϑ pour des estimations de Bayes, ceci ne signifie point qu'on ne peut pas trouver l'estimation du minimax.

En vertu de (2.77) on a pour $r(\vartheta) = r_{min}$

$$R_{max} = r_{min}, \tag{2.94}$$

c'est-à-dire que le maximum du risque moyen coïncide avec le minimum de la fonction de risque, si l'estimation du minimax est une estimation de Bayes.

2.4. ESTIMATIONS SIMULTANÉES D'UN ENSEMBLE DE PARAMÈTRES

2.4.1. Principales propriétés. La théorie des estimations exposée ci-dessus peut être étendue aux cas où la distribution initiale dépend de plusieurs paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$. A partir d'un échantillon de taille donnée n appartenant à la distribution $w_1(x; \vartheta_1, \ldots, \vartheta_m)$ on peut trouver m fonctions des valeurs échantillonnées (estimations conditionnelles)

$$\hat{\vartheta}_n^{(i)} = g^{(i)}(x_1, \ldots, x_n), \quad i = 1, \ldots, m, \tag{2.95}$$

en supposant que les valeurs des paramètres $\vartheta_1, \ldots, \vartheta_m$ sont données, ou encore m paramètres de la fonction

$$\hat{\Theta}_n^{(i)} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g^{(i)}(x_1, \ldots, x_n)\, w_m(\vartheta_1, \ldots, \vartheta_m)\, d\vartheta_1 \ldots d\vartheta_m \quad (2.95')$$

dont la moyenne est prise sur la distribution $w_m(\vartheta_1, \ldots, \vartheta_m)$ a priori conjointe (estimations inconditionnelles).

Toute estimation conditionnelle $\hat{\vartheta}_n^{(i)}$ est dite consistante si pour $n \to \infty$ elle converge en probabilité vers le paramètre ϑ_i et absolument correcte si pour un n quelconque sa moyenne sur l'ensemble des échantillons est égale à ϑ_i. D'une manière analogue chacune des estimations inconditionnelles est dite consistante si pour $n \to \infty$ elle converge en probabilité vers $m_1\{\vartheta_i\}$ et absolument correcte si pour tout n sa moyenne sur l'ensemble des échantillons est égale à $m_1\{\vartheta_i\}$.

Les estimations $g^{(i)}(x_1, \ldots, x_n)$, $i = 1, \ldots, m$ sont *simultanément suffisantes* si la fonction de vraisemblance de l'échantillon

$$L_x(\boldsymbol{\vartheta}) = W_n(x_1, \ldots, x_n \mid \vartheta_1, \ldots, \vartheta_m) = \prod_{k=1}^{n} w_1(x_k; \vartheta_1, \ldots, \vartheta_m), \quad (2.96)$$

$$\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_m),$$

peut s'écrire sous la forme d'un produit

$$L_x(\boldsymbol{\vartheta}) = f\,[g^{(1)}(x_1, \ldots, x_n), \ldots$$
$$\ldots, g^{(m)}(x_1, \ldots, x_n), \vartheta_1, \ldots, \vartheta_m]\, h(x_1, \ldots, x_n). \quad (2.97)$$

Soit $\hat{\vartheta}_n^{(i)} = g^{(i)}(x_1, \ldots, x_n)$ l'estimation non biaisée du paramètre ϑ_i, $i = 1, \ldots, m$. Considérons les moyennes suivantes prises sur l'ensemble des échantillons de taille donnée n:

$$I_n^{(i,j)}(\vartheta_1, \ldots, \vartheta_m) = m_1\left\{\frac{\partial}{\partial \vartheta_i} \ln L_x(\boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} \ln L_x(\boldsymbol{\vartheta})\right\} =$$
$$= -m_1\left\{\frac{\partial^2}{\partial \vartheta_i\, \partial \vartheta_j} \ln L_x(\boldsymbol{\vartheta})\right\}. \quad (2.98)$$

La matrice carrée de dimension $m \times m$ dont les éléments sont $I_n^{(i,j)}$ est appelée *matrice informationnelle de Fisher*. On voit à partir de (2.98) que la matrice informationnelle de Fisher est la matrice de corrélation de l'ensemble des variables aléatoires liées $\frac{\partial}{\partial \vartheta_i} \ln L_x(\boldsymbol{\vartheta})$, $i = 1, \ldots, m$, dont les moyennes sont nulles en vertu de (2.37).

Si le déterminant de la matrice informationnelle est différent de zéro, l'inégalité de Rao-Cramer (2.41) se trouve généralisée de la manière suivante. Pour tous $u_1, \ldots, u_m$ réels la forme quadra-

tique

$$\sum_{i=1}^{m}\sum_{j=1}^{m}[m_1\{(\hat{\vartheta}_n^{(i)}-\vartheta_i)(\hat{\vartheta}_n^{(j)}-\vartheta_j)\}-Y_n^{(i,\,j)}]\,u_iu_j\geqslant 0, \qquad (2.99)$$

où $Y_n^{(i,j)}$ est l'élément de la matrice inverse de la matrice informationnelle de Fisher, c'est-à-dire est égal au rapport du cofacteur de l'élément $I_n^{(i,j)}$ dans la matrice de Fisher au déterminant de cette matrice.

Le système d'estimations $\hat{\vartheta}_n^{(1)}, \ldots, \hat{\vartheta}_n^{(m)}$ pour lequel (2.99) devient l'égalité est dit *simultanément efficace*. Si ceci n'a lieu que pour $n \to \infty$ les estimations sont dites asymptotiquement simultanément efficaces.

Les éléments de la matrice informationnelle peuvent s'exprimer en fonction de la distribution initiale

$$I_n^{(i,\,j)}(\vartheta_1, \ldots, \vartheta_m)=nm_1\left\{\frac{\partial}{\partial\vartheta_i}\ln w_1(x\,;\,\vartheta_1, \ldots, \vartheta_m)\times\right.$$
$$\left.\times\frac{\partial}{\partial\vartheta_j}\ln w_1(x\,;\,\vartheta_1, \ldots, \vartheta_m)\right\}. \qquad (2.100)$$

Cette formule est analogue à (2.43) et coïncide avec cette dernière pour $m = 1$.

En utilisant la notion d'*ellipsoïde de corrélation* (voir annexe X) on peut donner une interprétation géométrique de l'inégalité (2.99): l'ellipsoïde de corrélation d'un système quelconque d'estimations des paramètres englobe toujours l'ellipsoïde décrit par l'équation de la forme

$$\sum_{i=1}^{m}\sum_{j=1}^{m}I_n^{(i,\,j)}z_iz_j=1. \qquad (2.101)$$

L'ellipsoïde de corrélation des estimations simultanément efficaces coïncide avec l'ellipsoïde (2.101). Le rapport du carré du volume de l'ellipsoïde (2.101) au carré du volume de l'ellipsoïde de corrélation (voir annexe X) peut être appelé *efficacité* des estimations simultanées d'un ensemble de paramètres.

Considérons maintenant les méthodes permettant d'obtenir les estimations d'un ensemble de paramètres à partir de critères de qualité donnés.

2.4.2. Méthode des moments. Bien que cette méthode soit basée sur des critères de qualité formulés plutôt d'une façon intuitive, elle est souvent utilisée par suite de sa simplicité.

Si la distribution initiale $w_1(x;\ \vartheta_1, \ldots, \vartheta_s)$ contient s paramètres inconnus et si le moment d'ordre $2s$ de cette distribution existe, en vertu de la loi des grands nombres tous les moments d'échantillon jusqu'à l'ordre s inclus, c'est-à-dire les grandeurs

$$m_k^*=\frac{1}{n}\sum_{i=1}^{n}x_i^k, \qquad k=1, \ldots, s,$$

convergent en probabilité, pour $n \to \infty$, vers les moments correspondants de la distribution initiale

$$m_k(\vartheta_1, \ldots, \vartheta_s) = \int_{-\infty}^{\infty} x^k w_1(x; \vartheta_1, \ldots, \vartheta_s)\, dx.$$

En égalant les moments d'échantillon et les moments de la distribution initiale on obtient un système de s équations par rapport aux paramètres inconnus

$$m_k(\vartheta_1, \ldots, \vartheta_s) = \frac{1}{n} \sum_{i=1}^{n} x_i^k, \qquad k = 1, \ldots, s, \tag{2.102}$$

dont les solutions

$$\hat{\vartheta}_n^{(k)} = g^{(k)}(x_1, \ldots, x_n), \quad k = 1, \ldots, s, \tag{2.103}$$

donnent les estimations par la méthode des moments. Pour $n \to \infty$ ces estimations sont asymptotiquement correctes, leur distribution étant asymptotiquement normale de variance décroissant comme $\frac{1}{n}$.

La méthode des moments peut être utilisée pour tout $s \gg 1$ fini.

2.4.3. Estimations du maximum de vraisemblance. Choisissons comme estimations (conditionnelles) des paramètres inconnus les valeurs $\vartheta_1, \ldots, \vartheta_m$, pour lesquelles pour des $x_1, \ldots, x_n$ donnés la fonction de vraisemblance (2.96), en tant que fonction de m variables ϑ_i, atteint un maximum. La condition d'extrémum s'écrit à l'aide du système d'équations

$$\frac{\partial L_x(\boldsymbol{\vartheta})}{\partial \vartheta_i} = 0, \qquad i = 1, \ldots, m. \tag{2.104}$$

Cependant, comme dans le cas de l'estimation d'un seul paramètre, à la place de (2.104) on utilise le système suivant d'équations de vraisemblance:

$$\frac{\partial \ln L_x(\boldsymbol{\vartheta})}{\partial \vartheta_i} = \sum_{k=1}^{n} \frac{\partial}{\partial \vartheta_i} \ln w_1(x_k; \vartheta_1, \ldots, \vartheta_m) = 0,$$
$$i = 1, \ldots, m. \tag{2.105}$$

Si pour les paramètres $\vartheta_1, \ldots, \vartheta_m$ les estimations non biaisées simultanément efficaces $\hat{\vartheta}_{n\,\mathrm{eff}}^{(1)}, \ldots, \hat{\vartheta}_{n\,\mathrm{eff}}^{(m)}$ existent, le système d'équations de vraisemblance a une solution *unique* égale à ces estimations.

Pour certaines restrictions concernant la fonction $w_1(x; \vartheta_1, \ldots, \vartheta_m)$ la solution du système d'équations de vraisemblance donne des estimations consistantes *), simultanément effi-

*) Si pour $n \to \infty$ le nombre m de paramètres inconnus croît avec la taille de l'échantillon, les estimations du maximum de vraisemblance peuvent ne pas être consistantes (cf. [2], § 45).

caces et asymptotiquement correctes, dont la distribution conjointe est asymptotiquement normale de moyennes égales aux paramètres estimés, et de matrice de corrélation qui est l'inverse de la matrice informationnelle de Fisher.

2.4.4. Maximum de densité de probabilité a posteriori des paramètres estimés. Supposons que les paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$ soient des variables aléatoires, dont la densité de probabilité conjointe est $w_m(\vartheta_1, \ldots, \vartheta_m)$. La densité de probabilité conditionnelle de ces paramètres, lorsque l'on extrait l'échantillon $x_1, \ldots, x_n$, est égale à

$$W_m(\vartheta_1, \ldots, \vartheta_m \mid x_1, \ldots, x_n) = \frac{w_m(\vartheta_1, \ldots, \vartheta_m) L_x(\boldsymbol{\vartheta})}{\int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} w_m(\vartheta_1, \ldots, \vartheta_m) L_x(\boldsymbol{\vartheta})\, d\vartheta_1 \ldots d\vartheta_m}. \quad (2.106)$$

Prenons pour estimations (non conditionnelles) des paramètres aléatoires les valeurs des variables $\vartheta_1, \ldots, \vartheta_m$ pour lesquelles, pour des $x_1, \ldots, x_n$ donnés, la densité de probabilité a posteriori (2.106) est maximale. Ces valeurs sont des solutions du système d'équations suivant [cf. (2.68) et (2.69)]:

$$\frac{\partial}{\partial \vartheta_i} \ln [w_m(\vartheta_1, \ldots, \vartheta_m) L_x(\boldsymbol{\vartheta})] = 0 \quad (2.107)$$

ou

$$\frac{\partial}{\partial \vartheta_i} \ln w_m(\vartheta_1, \ldots, \vartheta_m) + \sum_{k=1}^{n} \frac{\partial}{\partial \vartheta_i} \ln w_1(x_k \mid \vartheta_1, \ldots, \vartheta_m) = 0, \quad i = 1, \ldots, m. \quad (2.108)$$

Les estimations correspondant au maximum de densité de probabilité a posteriori ont des propriétés analogues à celles des estimations conditionnelles du maximum de vraisemblance examinées au paragraphe précédent. Si $\hat{\Theta}_u^{(1)} \ldots \hat{\Theta}_n^{(m)}$ sont les estimations inconditionnelles non biaisées simultanément efficaces des paramètres aléatoires $\vartheta_1, \ldots, \vartheta_m$, ces estimations sont l'unique solution du système d'équations (2.107).

Dans certains cas pour m paramètres inconnus, seule l'estimation de $s \leqslant m$ paramètres est intéressante, quant aux autres $m - s$ paramètres, il est préférable de ne pas les étudier du tout. Soient, par exemple, les m paramètres inconnus numérotés de telle sorte que les s paramètres qu'on veut estimer sont les premiers de la rangée. Dans ce cas par intégration sur les variables $\vartheta_{s+1}, \ldots, \vartheta_m$ on trouve, à partir de (2.106), la densité de probabilité

a posteriori de ces paramètres, soit:

$$W_s(\vartheta_1, \ldots, \vartheta_s \mid x_1, \ldots, x_n) =$$

$$= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} W_m(\vartheta_1, \ldots, \vartheta_s, \vartheta_{s+1}, \ldots, \vartheta_m \mid x_1, \ldots, x_n)\, d\vartheta_{s+1} \ldots d\vartheta_m,$$

puis pour obtenir les estimations simultanées $\hat{\vartheta}_n^{(1)}, \ldots, \hat{\vartheta}_n^{(s)}$ on maximise la fonction conditionnelle $W_s(\vartheta_1, \ldots, \vartheta_s \mid x_1, \ldots, x_n)$.

2.4.5. **Estimations de Bayes.** Lorsqu'on estime m paramètres inconnus, m erreurs apparaissent dotées chacune d'un certain poids. La fonction de pertes tenant compte des conséquences qu'entraînent ces erreurs dépend, dans le cas général, de $2m$ paramètres, soit:

$$\Pi = \Pi(\hat{\vartheta}_n^{(1)}, \ldots, \hat{\vartheta}_n^{(m)}, \vartheta_1, \ldots, \vartheta_m). \tag{2.109}$$

On prend parfois pour fonction de pertes la somme des fonctions de pertes d'un certain type pour quelques paramètres, sans tenir compte de la différence du poids des erreurs pour différents paramètres. Ainsi, on peut introduire des fonctions de pertes analogues à (2.71) à (2.74), notamment: la fonction de pertes simple

$$\Pi = mc - \sum_{k=1}^{m} \delta(\hat{\vartheta}_n^{(k)} - \vartheta_k); \tag{2.110}$$

la fonction de pertes, somme des modules des erreurs

$$\Pi = \sum_{k=1}^{m} \mid \hat{\vartheta}_n^{(k)} - \vartheta_k \mid; \tag{2.111}$$

et la fonction de pertes quadratique

$$\Pi = \sum_{k=1}^{m} (\hat{\vartheta}_n^{(k)} - \vartheta_k)^2. \tag{2.112}$$

Pour la fonction de pertes rectangulaire [voir (2.74)] on peut adopter la loi additive ou multiplicative de combinaison des composantes. On a alors

$$\Pi = \sum_{k=1}^{m} \Pi_k \tag{2.113}$$

ou

$$\Pi = \prod_{k=1}^{m} \Pi_k, \tag{2.113'}$$

ici

$$\Pi_k = \begin{cases} 1, & \mid \hat{\vartheta}_n^{(k)} - \vartheta_k \mid \geqslant \varepsilon, \\ 0, & \mid \hat{\vartheta}_n^{(k)} - \vartheta_k \mid < \varepsilon. \end{cases} \tag{2.113''}$$

Les estimations conditionnelles de Bayes des paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$ minimisent la fonction conditionnelle de risque

$$r(\vartheta_1, \ldots, \vartheta_m) = \\ = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \Pi(\hat{\vartheta}_n^{(1)}, \ldots, \hat{\vartheta}_n^{(m)}, \vartheta_1, \ldots, \vartheta_m) L_x(\boldsymbol{\vartheta}) \, dx_1 \ldots dx_n. \tag{2.114}$$

Si les paramètres estimés sont aléatoires, on prend pour critère de qualité le minimum du risque moyen, c'est-à-dire le minimum de la grandeur

$$R = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w_m(\vartheta_1, \ldots, \vartheta_m) \, r(\vartheta_1, \ldots, \vartheta_m) \, d\vartheta_1 \ldots d\vartheta_m, \tag{2.115}$$

qui peut également s'écrire comme suit:

$$R = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} J(\hat{\vartheta}_n^{(1)}, \ldots, \hat{\vartheta}_n^{(m)} \mid x_1, \ldots, x_n) \times \\ \times W_n(x_1, \ldots, x_n) \, dx_1 \ldots dx_n, \tag{2.116}$$

où

$$J(\hat{\vartheta}_n^{(1)}, \ldots, \hat{\vartheta}_n^{(m)} \mid x_1 \ldots, x_n) = \\ = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \Pi(\hat{\vartheta}_n^{(1)}, \ldots, \hat{\vartheta}_n^{(m)}, \vartheta_1, \ldots, \vartheta_m) \times \\ \times W_m(\vartheta_1, \ldots, \vartheta_m \mid x_1, \ldots, x_n) \, d\vartheta_1 \ldots d\vartheta_m, \tag{2.117}$$

$$W_n(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w_m(\vartheta_1, \ldots, \vartheta_m) L_x(\boldsymbol{\vartheta}) \, d\vartheta_1 \ldots d\vartheta_m. \tag{2.117'}$$

Les estimations satisfaisant au critère mentionné sont appelées *estimations inconditionnelles de Bayes des paramètres aléatoires* $\vartheta_1, \ldots, \vartheta_m$.

Le risque moyen R est une fonctionnelle multidimensionnelle dépendant des m fonctions $\hat{\vartheta}_n^{(i)} = g_i(x_1, \ldots, x_n)$, $i = 1, \ldots, n$. Comme ces dernières n'entrent que dans l'expression (2.117) et comme $W_n \geqslant 0$, le système d'équations donnant les fonctions extrémales de la fonctionnelle (2.116) peut s'écrire comme suit :

$$\frac{\partial J}{\partial g_i} = 0, \qquad i = 1, \ldots, m, \tag{2.118}$$

à condition toutefois que les dérivées dans (2.118) existent.

En généralisant (2.82) on remarque que les estimations conditionnelles du maximum de vraisemblance des paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$ sont une forme particulière des estimations conditionnelles de Bayes pour une fonction de pertes simple. D'une ma-

nière analogue, en généralisant (2.84) notons que les estimations inconditionnelles correspondant au maximum de densité de probabilité a posteriori des paramètres estimés sont un cas particulier des estimations inconditionnelles de Bayes pour une fonction de pertes simple.

Pour une fonction de pertes quadratique, en vertu de (2.112) et (2.118), les estimations de Bayes sont

$$\hat{\Theta}_n^{(i)*} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \vartheta_i W_m(\vartheta_1, \ldots, \vartheta_m | x_1, \ldots, x_n) \, d\vartheta_1 \ldots d\vartheta_m. \quad (2.119)$$

Après intégration sur toutes les variables, à l'exception de ϑ_i, la formule (2.119) devient

$$\hat{\Theta}_n^{(i)*} = \int_{-\infty}^{\infty} \vartheta_i W_1(\vartheta_i | x_1, \ldots, x_n) \, d\vartheta_i. \quad (2.120)$$

Ainsi, pour une fonction de pertes quadratique, on peut trouver séparément l'estimation de Bayes de chacun des paramètres liés $\vartheta_1, \ldots, \vartheta_m$, comme dans le cas de l'estimation d'un seul paramètre [cf. (2.87)].

Les résultats du § 2.3 relatifs à d'autres fonctions de pertes (cf. [9], § 21.2.3) peuvent également être étendus au cas de l'estimation d'un ensemble d'un nombre arbitraire fini de paramètres. Enfin, il n'y a aucune difficulté de principe à généraliser la notion d'estimations du minimax.

2.5. ESTIMATIONS DES PARAMÈTRES DE CERTAINES DISTRIBUTIONS UNIDIMENSIONNELLES

2.5.1. Estimation conditionnelle du paramètre de la loi exponentielle. A titre d'exemple d'une loi uniparamétrique nous allons étudier la loi exponentielle. La densité de probabilité d'une variable aléatoire distribuée suivant cette loi est [cf. (1.95)]

$$w_1(x;\ \vartheta) = \frac{1}{\vartheta} e^{-\frac{x}{\vartheta}},$$
$$x > 0,\ \vartheta > 0. \quad (2.121)$$

La moyenne de cette variable aléatoire est égale à ϑ et sa variance est ϑ^2 (cf. problème 2.1, tome I).

La fonction de vraisemblance d'un échantillon de taille n, composé d'éléments indépendants, est en vertu de (2.10)

$$W_n(x_1, \ldots, x_n | \vartheta) = \frac{1}{\vartheta^n} \exp\left(-\frac{1}{\vartheta} \sum_{i=1}^{n} x_i\right). \quad (2.122)$$

En comparant (2.122) et (2.28), on voit tout de suite que pour estimer le paramètre inconnu ϑ de la loi exponentielle, on n'a pas besoin de connaître chaque valeur échantillonnée séparément, il suffit de connaître leur somme

$$g(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i, \tag{2.123}$$

qui dans le cas général est une statistique suffisante. Comme la variance de la loi exponentielle est limitée, la moyenne échantillonnée $\frac{1}{n} \sum_{i=1}^{n} x_i$ en vertu du résultat général (2.23) est une estimation consistante et absolument correcte de la moyenne ϑ de la loi exponentielle.

Pour illustrer les assertions du § 2.2.4 notons que la statistique suffisante (2.123) impose la relation suivante pour le rapport de vraisemblance :

$$l(x_1, \ldots, x_n \mid \vartheta) = \left(\frac{\vartheta_0}{\vartheta}\right)^n \exp\left[-\left(\frac{1}{\vartheta} - \frac{1}{\vartheta_0}\right) \sum_{i=1}^{n} x_i\right].$$

En utilisant (2.43), on peut trouver l'information de Fisher pour la loi exponentielle, soit :

$$I_n(\vartheta) = n \int_0^{\infty} \left(\frac{1}{\vartheta} - \frac{x}{\vartheta^2}\right) \frac{1}{\vartheta} e^{-\frac{x}{\vartheta}} dx =$$

$$= \frac{n}{\vartheta^4} \int_0^{\infty} (\vartheta - x)^2 \frac{1}{\vartheta} e^{-\frac{x}{\vartheta}} dx = \frac{n}{\vartheta^2}. \tag{2.124}$$

De plus, la variance de la moyenne échantillonnée est toujours (lorsque les éléments de l'échantillon de taille n sont indépendants) n fois inférieure à la variance de la loi initiale, c'est-à-dire

$$M_2 \left\{\frac{1}{n} \sum_{i=1}^{n} x_i\right\} = \frac{\vartheta^2}{n}. \tag{2.125}$$

En vertu de (2.124) et (2.125) on a

$$M_2 \left\{\frac{1}{n} \sum_{i=1}^{n} x_i\right\} = \frac{1}{I_n(\vartheta)}. \tag{2.126}$$

Ainsi, la moyenne échantillonnée

$$\hat{\vartheta}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.127}$$

est une estimation efficace absolument correcte du paramètre ϑ de la loi exponentielle (2.121).

Notons que la dérivée du logarithme de la fonction de vraisemblance (2.122)

$$\frac{\partial}{\partial\vartheta}\ln W_n(x_1, \ldots, x_n \mid \vartheta) = -\frac{n}{\vartheta} + \frac{1}{\vartheta^2}\sum_{i=1}^{n} x_i =$$

$$= \frac{n}{\vartheta^2}(\hat{\vartheta}_n - \vartheta) \qquad (2.128)$$

coïncide exactement avec la représentation (2.46), compte tenu de (2.47″).

Il est facile de voir maintenant que la moyenne échantillonnée (2.127) est la solution unique de l'équation de vraisemblance

$$\frac{\partial}{\partial\vartheta}\ln W_n(x_1, \ldots, x_n \mid \vartheta) = 0,$$

ce qui est en accord avec le résultat général du § 2.3.1.

Comme la variable aléatoire $\frac{2}{\vartheta}\sum_{k=1}^{n} x_k$ est répartie suivant la loi du χ^2 à $2n$ degrés de liberté (cf. § 1.3.3), la variable $\frac{2n}{\vartheta}\sum_{k=1}^{n} x_k$ est répartie suivant cette même loi. L'estimation d'intervalle [cf. (2.50)] du paramètre de la loi exponentielle peut alors être calculée à partir de la relation

$$P\left\{2n(1-\varepsilon_2) < \frac{2n\hat{\vartheta}_n}{\vartheta} < 2n(1+\varepsilon_1)\right\} = \gamma. \qquad (2.129)$$

Pour calculer ε_1 et ε_2 nous allons utiliser (2.51) et (2.51′) en posant $\gamma_1 = \gamma_2 = \frac{1}{2} - \frac{\gamma}{2}$. On a alors

$$1+\varepsilon_1 = \frac{1}{2n}\chi^2_{\frac{1}{2}-\frac{\gamma}{2}}, \qquad (2.130)$$

$$1-\varepsilon_2 = \frac{1}{2n}\chi^2_{\frac{1}{2}+\frac{\gamma}{2}}, \qquad (2.131)$$

où χ^2_δ est le quantile d'une variable aléatoire répartie suivant la loi du χ^2 à $2n$ degrés de liberté.

2.5.2. Estimations de Bayes du paramètre de la loi exponentielle. Supposons que le paramètre ϑ d'une loi exponentielle soit une variable aléatoire, répartie également suivant une loi exponentielle de paramètre ϑ_0 donné, soit :

$$w_1(\vartheta) = \frac{1}{\vartheta_0}e^{-\frac{\vartheta}{\vartheta_0}}, \quad \vartheta > 0, \ \vartheta_0 > 0. \qquad (2.132)$$

Nous commençons par l'estimation de Bayes du paramètre pour une fonction de pertes simple, qui est également une estimation inconditionnelle correspondant au maximum de densité de probabilité a posteriori du paramètre estimé. Cette estimation est donnée par l'équation (2.69) qui, compte tenu de (2.122) et (2.132), est

$$-\frac{1}{\vartheta_0}+\frac{1}{\vartheta^2}\sum_{k=1}^{n} x_k-\frac{n}{\vartheta}=0. \tag{2.133}$$

L'estimation cherchée peut être obtenue à partir de (2.133)

$$\hat{\vartheta}_n=\frac{n\vartheta_0}{2}\left(\sqrt{1+\frac{4}{n\vartheta_0}\cdot\frac{1}{n}\sum_{k=1}^{n} x_k}-1\right). \tag{2.134}$$

Lorsque la taille de l'échantillon augmente indéfiniment $(n\to\infty)$, on a

$$\hat{\vartheta}_n\sim\frac{1}{n}\sum_{k=1}^{n} x_k. \tag{2.134'}$$

Ainsi, la moyenne arithmétique est asymptotiquement une estimation de Bayes. Pour un échantillon de taille limitée l'algorithme (2.134) donnant l'estimation de Bayes d'après les valeurs échantillonnées est *non linéaire.*

On arrive également à l'estimation du maximum de vraisemblance (2.134') dans le cas où $\vartheta_0\to\infty$, ce qui correspond à peu près à une répartition uniforme du paramètre ϑ.

Calculons également l'estimation de Bayes pour une fonction de pertes quadratique en utilisant (2.87). Comme dans le cas envisagé on a

$$W_1(\vartheta\mid x_1,\ \ldots,\ x_n)=\frac{\dfrac{1}{\vartheta_0\vartheta^n}\exp\left\{-\left(\dfrac{\vartheta}{\vartheta_0}+\dfrac{1}{\vartheta}\displaystyle\sum_{k=1}^{n} x_k\right)\right\}}{\dfrac{1}{\vartheta_0}\displaystyle\int_0^\infty\dfrac{1}{\vartheta^n}\exp\left\{-\left(\dfrac{\vartheta}{\vartheta_0}+\dfrac{1}{\vartheta}\displaystyle\sum_{k=1}^{n} x_k\right)\right\}d\vartheta},$$

l'estimation de Bayes cherchée, égale à la moyenne conditionnelle, est

$$\hat{\vartheta}_n=\frac{\displaystyle\int_0^\infty\frac{1}{\vartheta^{n-1}}\exp\left\{-\left(\frac{\vartheta}{\vartheta_0}+\frac{1}{\vartheta}\sum_{k=1}^{n} x_k\right)\right\}d\vartheta}{\displaystyle\int_0^\infty\frac{1}{\vartheta^{n}}\exp\left\{-\left(\frac{\vartheta}{\vartheta_0}+\frac{1}{\vartheta}\sum_{k=1}^{n} x_k\right)\right\}d\vartheta}. \tag{2.135}$$

Comme

$$\int_0^\infty x^{-n} e^{-\left(x+\frac{a^2}{x}\right)} dx = \frac{2}{a^{n-1}} K_{n-1}(2a), \qquad (2.136)$$

où $K_m(z)$ est une fonction de Bessel de seconde espèce de l'argument imaginaire, en vertu de (2.135) on a

$$\hat{\vartheta}_n = \vartheta_0 \sqrt{\frac{1}{\vartheta_0}\sum_{k=1}^n x_k}\, \frac{K_{n-2}\left(2\sqrt{\frac{1}{\vartheta_0}\sum_{k=1}^n x_k}\right)}{K_{n-1}\left(2\sqrt{\frac{1}{\vartheta_0}\sum_{k=1}^n x_k}\right)}. \qquad (2.137)$$

En comparant (2.137) et (2.134) on voit que les estimations de Bayes pour des fonctions de pertes simple et quadratique sont essentiellement différentes, comme on pouvait s'y attendre car la densité de probabilité conditionnelle (2.132) *n'est pas symétrique.*

2.5.3. Estimations conditionnelles des paramètres de la loi normale. A titre d'exemple d'estimation simultanée de plusieurs paramètres d'une distribution unidimensionnelle nous allons envisager les estimations de la moyenne a et de la variance $\sigma^2 < \infty$ de la loi normale. La fonction de vraisemblance d'un échantillon de taille n dont les éléments sont indépendants est, en vertu de (2.10), pour la loi mentionnée

$$L_x(a, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - a)^2\right] =$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n x_i^2 - 2a\sum_{i=1}^n x_i + na^2\right]\right\}. \qquad (2.138)$$

En comparant (2.138) et (2.97) on voit que $\sum_{i=1}^n x_i$ et $\sum_{i=1}^n x_i^2$ sont des statistiques simultanément suffisantes des paramètres de la loi normale. Comme la variance est limitée et vu les relations générales (2.23) et (2.25), vraies pour toute loi de distribution, la moyenne échantillonnée

$$\hat{a} = g_1(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i \qquad (2.139)$$

et la variance échantillonnée multipliée par $\frac{n}{n-1}$

$$\hat{\sigma}^2 = g_2(x_1, \ldots, x_n) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \hat{a})^2 \qquad (2.140)$$

sont les estimations consistantes et absolument correctes respectivement de la moyenne et de la variance de la loi normale.

Calculons maintenant les éléments de la matrice informationnelle de Fisher. En vertu de (2.98) et compte tenu de (2.138) on obtient :

$$I_n^{(1,1)}(a, \sigma^2) = m_1 \left\{ \left[\frac{\partial}{\partial a} \ln L_x(a, \sigma^2) \right]^2 \right\} = \\ = m_1 \left\{ \frac{1}{\sigma^4} \left[\sum_{i=1}^{n} (x_i - a) \right]^2 \right\} = \frac{n}{\sigma^2},$$

$$I_n^{(1,2)}(a, \sigma^2) = I^{(2,1)}(a, \sigma^2) = m_1 \left\{ \frac{\partial}{\partial a} \ln L_x(a, \sigma^2) \frac{\partial}{\partial \sigma^2} \ln L_x(a, \sigma^2) \right\} = \\ = m_1 \left\{ \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - a) \left[\frac{1}{2\sigma^4} \sum_{k=1}^{n} (x_k - a)^2 - \frac{n}{2\sigma^2} \right] \right\} = 0,$$

$$I_n^{(2,2)}(a, \sigma^2) = m_1 \left\{ \left[\frac{\partial}{\partial \sigma^2} \ln L_x(a, \sigma^2) \right]^2 \right\} = \\ = m_1 \left\{ \left[\frac{1}{2\sigma^4} \sum_{k=1}^{n} (x_k - a)^2 - \frac{n}{2\sigma^2} \right]^2 \right\} = \frac{n}{2\sigma^4}.$$

La matrice informationnelle est donc

$$\mathbf{I} = \left\| \begin{matrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{matrix} \right\|. \tag{2.141}$$

Le déterminant de cette matrice est

$$\det \mathbf{I} = \frac{n^2}{2\sigma^6} > 0,$$

et les éléments de la matrice inverse de la matrice informationnelle

$$Y_n^{(1,1)} = \frac{\sigma^2}{n}, \quad Y_n^{(2,2)} = \frac{2\sigma^4}{n},$$

$$Y_n^{(1,2)} = Y_n^{(2,1)} = 0.$$

D'un autre côté, les variances et la covariance des estimations (2.139) et (2.140) sont

$$M_2\{\hat{a}\} = \frac{\sigma^2}{n},$$

$$M_2\{\hat{\sigma}^2\} = m_1 \left\{ \left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{a})^2 - \sigma^2 \right]^2 \right\} = \frac{2\sigma^4}{n-1},$$

$$m_1\{(\hat{a} - a)(\hat{\sigma}^2 - \sigma^2)\} = \\ = \frac{1}{n-1} m_1 \left\{ (\hat{a} - a) \left[\sum_{i=1}^{n} (x_i - a)^2 - n(\hat{a} - a)^2 \right] \right\} = 0.$$

En portant dans (2.99) les expressions obtenues on trouve

$$\frac{2\sigma^4}{n(n-1)}u_2^2>0,$$

c'est-à-dire que les estimations (2.139) et (2.140) ne sont pas *simultanément* efficaces. L'ellipsoïde de corrélation de ces estimations qui est donné par l'équation

$$\frac{z_1^2}{\sigma^2/n}+\frac{z_2^2}{2\sigma^4/(n-1)}=1,$$

ne coïncide pas mais englobe l'ellipsoïde informationnel [cf.(2.101)]

$$\frac{z_1^2}{\sigma^2/n}+\frac{z_2^2}{2\sigma^4/n}=1.$$

Calculons maintenant les estimations du maximum de vraisemblance. En portant (2.138) dans (2.105) on obtient le système d'équations de vraisemblance

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k-a)=0, \tag{2.142}$$

$$\frac{1}{2\sigma^4}\sum_{k=1}^{n}(x_k-a)^2-\frac{n}{2\sigma^2}=0. \tag{2.142'}$$

En vertu de la première équation, on a

$$\hat{a}=\frac{1}{n}\sum_{k=1}^{n}x_k, \tag{2.143}$$

c'est-à-dire que l'estimation du maximum de vraisemblance de la moyenne d'une variable aléatoire normale est égale à la moyenne échantillonnée. Notons que l'équation (2.142) ne dépend pas en fait du second paramètre σ^2 de la loi normale.

En comparant $M_2\{\hat{a}\}$ et $I_n^{(1,1)}$ on voit que la moyenne échantillonnée est une estimation efficace non biaisée de la moyenne d'une variable aléatoire normale.

En portant dans l'équation (2.142′) $\hat{a}$ au lieu de a on trouve l'estimation du maximum de vraisemblance de la variance:

$$\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\hat{a})^2=M_2^*. \tag{2.144}$$

Ainsi, l'estimation du maximum de vraisemblance de la variance d'une variable aléatoire normale est égale à la variance échantillonnée. Cette estimation est consistante et biaisée, le biais de σ^2 étant en vertu de (2.84) égal à

$$b_n(\sigma^2)=-\frac{\sigma^2}{n}. \tag{2.145}$$

La variance de l'estimation (2.144) est

$$M_2\{M_2^*\}=\frac{2\sigma^4}{n}\left(1-\frac{1}{n}\right). \tag{2.146}$$

L'infimum de la variance des estimations du paramètre σ^2 d'une variable aléatoire normale est, en vertu de l'inégalité de Rao-Cramer, égal à

$$M_2\{\hat{\sigma}^2\}_{\min}=\frac{2\sigma^4\left(1-\frac{1}{n}\right)^2}{n}. \tag{2.146'}$$

Il découle de (2.146) et (2.146') que l'efficacité de l'estimation du maximum de vraisemblance de la variance d'une variable aléatoire normale est égale à

$$l=\frac{M_2\{\hat{\sigma}^2\}_{\min}}{M_2\{M_2^*\}}=1-\frac{1}{n} \tag{2.147}$$

Ainsi, l'estimation (2.144) n'est pas efficace, mais seulement asymptotiquement efficace en vertu des propriétés générales des estimations du maximum de vraisemblance mentionnées ci-dessus.

Notons que si l'on connaît la moyenne a priori, l'estimation du maximum de vraisemblance de la variance d'une variable aléatoire normale est [cf. (2.142')]

$$\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-a)^2. \tag{2.148}$$

Cette estimation, comme nous l'avons noté au § 2.2.3, est absolument correcte, sa variance étant

$$M_2\{\hat{\sigma}^2\}=\frac{2\sigma^4}{n}, \tag{2.148'}$$

ce qui est l'inverse de l'information $I_n^{(2,2)}$. Ainsi, si l'on connaît la moyenne, l'estimation du maximum de vraisemblance de la variance d'une variable aléatoire normale est efficace.

Jusqu'à présent nous avons envisagé des estimations ponctuelles (conditionnelles) des paramètres de la loi normale. Examinons maintenant les estimations d'intervalle de ces paramètres.

Commençons par l'estimation d'intervalle de la moyenne d'une variable aléatoire normale, ceci en supposant que le second paramètre σ^2 soit exactement connu. Pour l'estimation ponctuelle de la moyenne prenons la moyenne échantillonnée (2.139). Introduisons l'erreur normée de l'estimation

$$\varepsilon=\frac{a-\hat{a}}{\sqrt{\frac{\sigma^2}{n}}} \tag{2.149}$$

et calculons la probabilité pour que la valeur absolue de cette erreur soit inférieure à une valeur donnée ε_γ

$$P\{|\varepsilon| < \varepsilon_\gamma\} = \gamma. \tag{2.150}$$

A partir de (2.149) on voit immédiatement que l'erreur normée de l'estimation de la moyenne est une variable aléatoire normale de moyenne nulle et de variance unité. On peut alors écrire (2.150) sous la forme suivante

$$2F(\varepsilon_\gamma) - 1 = \gamma \tag{2.151}$$

ou

$$\varepsilon_\gamma = x_{\frac{1}{2}-\frac{\gamma}{2}}, \tag{2.151'}$$

où $F(x)$ est l'intégrale de Laplace, et $x_{\frac{1}{2}-\frac{\gamma}{2}}$ le quantile de la loi normale (voir *annexe* VII).

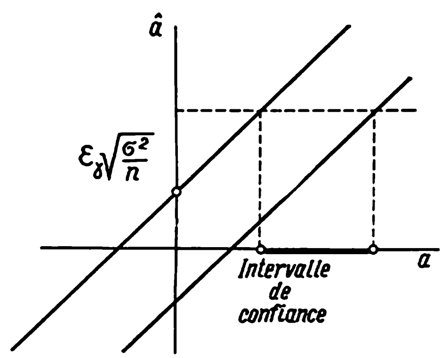


Fig. 2.5. Construction d'un intervalle de confiance

Les deux dernières équations expriment la relation entre la longueur de l'intervalle de confiance $2\varepsilon_\gamma$ pour l'erreur normée et le coefficient de confiance γ. La première est utilisée pour le calcul de γ si ε_γ est donné, et la seconde, pour le calcul de ε_γ lorsque γ est donné. Pour le paramètre estimé a, l'intervalle de confiance peut être donné par les inégalités suivantes [cf. (2.149) et (2.150)]:

$$\hat{a} - \varepsilon_\gamma\sqrt{\frac{\sigma^2}{n}} < a < \hat{a} + \varepsilon_\gamma\sqrt{\frac{\sigma^2}{n}}. \tag{2.152}$$

La relation existant entre $\hat{a}$ et a est géométriquement représentée par deux droites parallèles à la bissectrice de l'angle des coordonnées qui coupent sur l'axe $\hat{a}$ des portions $\pm\varepsilon_\gamma\sqrt{\frac{\sigma^2}{n}}$ (fig. 2.5). Pour trouver les bornes inférieure et supérieure de l'intervalle de confiance il faut projeter sur l'axe des abscisses les points d'intersection de ces droites avec la droite $\hat{a} = \text{const}$.

Considérons l'estimation d'intervalle de la moyenne d'une variable aléatoire normale lorsqu'on ne connaît pas sa variance. Utilisons en qualité d'estimations ponctuelles de la moyenne et de la variance les estimations non biaisées (2.139) et (2.140). Comme dans le cas précédent, on introduit ici l'erreur normée de l'estimation. La différence consiste en ce que pour la normalisation on utilise ici l'estimation $\hat{\sigma}^2$, la valeur exacte de σ^2 étant inconnue. Pour l'erreur normée de l'estimation de la moyenne on prend ainsi

la grandeur *)

$$t=\frac{\hat{a}-a}{\sqrt{\frac{\hat{\sigma}^2}{n}}}, \qquad (2.153)$$

qui est égale au rapport de deux variables aléatoires indépendantes, dont l'une est la variable normale $(\hat{a}-a)\left(\frac{\sigma^2}{n}\right)^{-1/2}$ (de moyenne nulle et de variance unité) et l'autre la variable aléatoire $\left(\frac{\hat{\sigma}^2}{\sigma^2}\right)^{1/2}$ répartie comme $\sqrt{\frac{\chi^2}{n-1}}$ (χ^2 ayant $n-1$ degrés de liberté). La distribution de ce rapport coïncide avec la *loi de Student* à $n-1$ degrés de liberté. Ainsi, la densité de probabilité de l'erreur normée s'écrit comme suit:

$$s_{n-1}(x)=\frac{1}{\sqrt{\pi(n-1)}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(1+\frac{x^2}{n-1}\right)^{-\frac{n}{2}}. \qquad (2.154)$$

Avec l'augmentation de la taille n de l'échantillon les deux premiers facteurs dans (2.154) tendent vers $(2\pi)^{-1/2}$, et le dernier vers

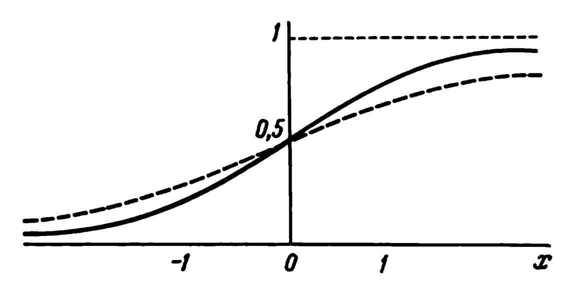


Fig. 2.6. Lois de distribution:
— normale; — — — de Student pour $n=2$

$e^{-x^2/2}$, de telle sorte que la fonction $s_n(x)$ pour $n\to\infty$ tend asymptotiquement vers la densité de l'écart normé de la variable aléatoire normale

$$s_n(x)\sim\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \qquad (2.154')$$

Ainsi, pour un échantillon de taille importante on peut, même si l'on ne connaît pas la variance, utiliser les formules (2.151), (2.151'), comme on pouvait s'y attendre ayant en vue la consistance de l'estimation $\hat{\sigma}^2$. Cependant, pour des n peu importants, la loi de Student diffère notablement de la loi normale (fig. 2.6).

*) On a utilisé la désignation t, ce qui est de tradition en statistique.

Compte tenu de la symétrie de la loi de Student, la probabilité pour que la valeur absolue de l'erreur normée ne soit pas supérieure à la grandeur t_γ est

$$P\{|t|<t_\gamma\}=2\int_0^{t_\gamma} s_{n-1}(x)\,dx=\gamma. \qquad (2.155)$$

Dans l'annexe IX on donne la table des nombres t_γ satisfaisant à la relation (2.155) pour des valeurs de n et γ variant dans une large gamme. La dernière ligne de la table correspond à la loi normale ($n = \infty$) et la valeur t_γ (et, donc, la longueur de l'intervalle de confiance) dans cette ligne est la plus petite pour tout γ donné. Un intervalle de confiance plus large que dans le cas précédent (cas où la variance était connue) obtenu pour les mêmes tailles de l'échantillon et le coefficient de confiance est le coût d'une information incomplète sur la variance σ^2 de l'estimation de la moyenne a.

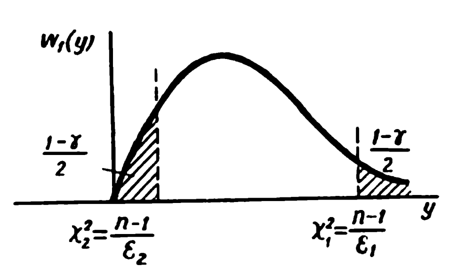


Fig. 2.7. Détermination des limites de confiance pour la variance d'une variable aléatoire normale

L'intervalle de confiance pour le paramètre estimé est maintenant donné par les inégalités suivantes [cf. (2.153) et (2.155)]:

$$\hat{a}-t_\gamma\sqrt{\frac{\hat{\sigma}^2}{n}}<a<\hat{a}+t_\gamma\sqrt{\frac{\hat{\sigma}^2}{n}}. \qquad (2.156)$$

Considérons en conclusion la construction de l'intervalle de confiance pour une variance inconnue σ^2 d'une variable aléatoire. Prenons pour l'estimation ponctuelle de la variance l'estimation non biaisée (2.140). La variable aléatoire

$$\frac{n-1}{\sigma^2}\hat{\sigma}^2=\sum_{k=1}^{n}\left(\frac{x_i-\hat{a}}{\sigma}\right)^2 \qquad (2.157)$$

est répartie suivant la loi du χ^2 à $n-1$ degrés de liberté. Prenons les grandeurs $\varepsilon_1\hat{\sigma}^2$ et $\varepsilon_2\hat{\sigma}^2$ pour les bornes inférieure et supérieure de confiance. La probabilité pour que l'intervalle de confiance avec les bornes indiquées contienne le paramètre σ^2 est égale à

$$P\{\varepsilon_1\hat{\sigma}^2<\sigma^2<\varepsilon_2\hat{\sigma}^2\}=P\left\{\frac{n-1}{\varepsilon_2}<\frac{n-1}{\sigma^2}\hat{\sigma}^2<\frac{n-1}{\varepsilon_1}\right\}=$$
$$=P\left\{\frac{n-1}{\varepsilon_2}<\chi^2<\frac{n-1}{\varepsilon_1}\right\}=\gamma. \qquad (2.158)$$

Les grandeurs $\varepsilon_1 = \frac{n-1}{\chi_1^2}$ et $\varepsilon_2' = \frac{n-1}{\chi_2^2}$ sont données par la condition (fig. 2.7)

$$P\{\chi^2 < \chi_2^2\} = P\{\chi^2 > \chi_1^2\} = \frac{1-\gamma}{2}, \tag{2.159}$$

ce qui équivaut à

$$\chi_1^2 = \chi^2_{\frac{1}{2}-\frac{\gamma}{2}}, \quad \chi_2^2 = \chi^2_{\frac{1}{2}+\frac{\gamma}{2}}. \tag{2.159'}$$

Ainsi, l'intervalle de confiance pour la variance d'une variable aléatoire normale, correspondant au coefficient de confiance γ, est déterminé par les inégalités suivantes:

$$\frac{n-1}{\chi^2_{\frac{1}{2}-\frac{\gamma}{2}}}\,\hat{\sigma}^2 < \sigma^2 < \frac{n-1}{\chi^2_{\frac{1}{2}+\frac{\gamma}{2}}}\,\hat{\sigma}^2. \tag{2.160}$$

Dans l'annexe VIII on trouvera la table des quantiles χ_δ^2 de la loi du khi carré.

2.5.4. Méthode des moindres carrés. Supposons que l'on mesure une certaine grandeur physique x dépendant d'une autre grandeur z, et supposons que l'on connaisse la relation fonctionnelle $x = f(z;\ a_1, a_2, \ldots)$ à l'exception de certains paramètres $a_1, a_2, \ldots$ Supposons que les résultats des mesures $x_1, \ldots, x_n$ soient des variables aléatoires *normales* indépendantes, dont les moyennes sont $f(z_1;\ a_1, a_2, \ldots), \ldots, f(z_n;\ a_1, a_2, \ldots)$ et les variances sont égales entre elles et égales à σ^2 (mesures équiprécises, absolument correctes). Pour déterminer *complètement* les relations fonctionnelles entre les grandeurs physiques x et z il y a lieu d'utiliser les résultats des mesures. La densité de probabilité conjointe des résultats des mesures (fonction de vraisemblance) est égale à

$$L_x(a_1, a_2, \ldots) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \times \right.$$
$$\left. \times \sum_{k=1}^{n} [x_k - f(z_k;\ a_1, a_2, \ldots)]^2 \right\}. \tag{2.161}$$

Nous allons exiger que les valeurs des paramètres inconnus $a_1, a_2, \ldots$, pour des résultats des mesures donnés, maximisent la fonction (2.161), ce qui équivaut à la condition de minimum de la grandeur

$$\sum_{k=1}^{n} [x_k - f(z_k;\ a_1, a_2, \ldots)]^2,$$

cette condition étant utilisée dans la *méthode* classique des *moindres carrés*.

Les valeurs des paramètres a_1, a_2, . . . sont données par le système d'équations

$$\sum_{k=1}^{n} [x_k - f(z_k;\ a_1,\ a_2,\ \ldots)] \frac{\partial f(z_k;\ a_1,\ a_2,\ \ldots)}{\partial a_j} = 0,\ j = 1,\ 2,\ \ldots \tag{2.162}$$

Si la relation existant entre les grandeurs physiques x et z est *linéaire*, c'est-à-dire si

$$x = a_1 z + a_2, \tag{2.163}$$

on a

$$\frac{\partial f(z_k;\ a_1,\ a_2)}{\partial a_1} = z_k, \quad \frac{\partial f(z_k;\ a_1,\ a_2)}{\partial a_2} = 1, \tag{2.163'}$$

et le système d'équations (2.162) s'écrit alors comme suit:

$$\sum_{k=1}^{n} (x_k - a_1 z_k - a_2)\, z_k = 0.$$

$$\sum_{k=1}^{n} (x_k - a_1 z_k - a_2) = 0.$$

Il en découle

$$\hat{a}_2 = m_1^* \{x\} - \hat{a}_1 \langle z \rangle, \tag{2.164}$$

$$\hat{a}_1 = \frac{m_1^* \{xz\} - m_1^* \{x\} \langle z \rangle}{\langle z^2 \rangle - \langle z \rangle^2}, \tag{2.164'}$$

où $m_1^* \{x\}$, $m_1^* \{xz\}$ sont les moyennes échantillonnées des variables aléatoires x et xz et

$$\langle z \rangle = \frac{1}{n} \sum_{i=1}^{n} z_i, \quad \langle z^2 \rangle = \frac{1}{n} \sum_{i=1}^{n} z_i^2.$$

Ainsi, la méthode des moindres carrés donne la relation linéaire suivante entre les grandeurs physiques x et z d'après les données des mesures:

$$x - m_1^* \{x\} = - \frac{m_1^* \{xz\} - m_1^* \{x\} \langle z \rangle}{\langle z^2 \rangle - \langle z \rangle^2} (z - \langle z \rangle). \tag{2.165}$$

La formule (2.165) est l'équation empirique de la droite de la régression quadratique moyenne [cf. (4) dans l'annexe XI].

2.5.5. Estimations de Bayes des paramètres de la loi normale. Nous allons commencer par le cas où la variance σ^2 de la loi normale est connue exactement, et la moyenne est une variable aléatoire normale de paramètres $(a_0,\ \sigma_0^2)$, c'est-à-dire

$$w_1(a) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{(a - a_0)^2}{2\sigma_0^2}}. \tag{2.166}$$

Calculons l'estimation correspondant au maximum de densité de probabilité a posteriori du paramètre estimé a laquelle en vertu de (2.84) est un cas particulier de l'estimation de Bayes pour une fonction de pertes simple. A partir de (2.69) on obtient l'équation donnant l'estimation cherchée

$$-\frac{a-a_0}{\sigma_0^2}+\frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k-a)=0,$$

d'où

$$\hat{a}=\frac{1}{1+\frac{\sigma^2}{n\sigma_0^2}}\left(\frac{1}{n}\sum_{k=1}^{n}x_k+\frac{a_0\sigma^2}{n\sigma_0^2}\right). \tag{2.167}$$

L'estimation (2.167) est la moyenne pondérée des deux grandeurs suivantes: de l'estimation du maximum de vraisemblance $\frac{1}{n}\sum_{k=1}^{n}x_k$ et de la moyenne a priori a_0 du paramètre estimé, le rapport du poids attribué à la seconde grandeur à celui de la première étant égal à $\frac{\sigma^2}{n\sigma_0^2}$, c'est-à-dire au rapport de la variance de l'estimation du maximum de vraisemblance à la variance a priori.

La densité de probabilité a posteriori du paramètre a pour un échantillon donné est

$$W_1(a\,|\,x_1,\ \ldots,\ x_n)=\sqrt{\frac{n}{2\pi\sigma^2}}\left(1+\frac{\sigma^2}{n\sigma_0^2}\right)^{1/2}\times$$
$$\times\exp\left\{-\frac{n}{2\sigma^2}\left(1+\frac{\sigma^2}{n\sigma_0^2}\right)\times\right.$$
$$\left.\times\left[a-\frac{1}{1+\frac{\sigma^2}{n\sigma_0^2}}\left(\frac{1}{n}\sum_{k=1}^{n}x_k+\frac{a_0\sigma^2}{n\sigma_0^2}\right)\right]^2\right\}. \tag{2.168}$$

Pour $n\to\infty$ la densité de probabilité (2.168) tend vers la fonction delta $\delta\left(a-\frac{1}{n}\sum_{k=1}^{n}x_k\right)$. Pour un n donné et $\left(\frac{\sigma_0}{\sigma}\right)^2\to\infty$ les paramètres de la fonction (2.168) tendent vers $\left(\frac{1}{n}\sum_{k=1}^{n}x_k,\ \frac{\sigma^2}{n}\right)$ et pour $\frac{\sigma^2}{n}\gg\sigma_0^2$ (2.168) devient (2.166).

On voit également que le second membre de (2.167) coïncide exactement avec la moyenne conditionnelle du paramètre estimé, c'est-à-dire avec la grandeur $m_1\{a\mid x_1,\ \ldots,\ x_n\}$. Par conséquent, (2.167) est également une estimation de Bayes pour une fonction de

pertes *quadratique*, ce qui correspond au résultat général mentionné au § 2.3.6 (car la distribution (2.166) est symétrique par rapport à son mode unique $a = a_0$). Le § 2.3.8 permet de conclure que l'estimation en question est une estimation de Bayes pour une fonction de pertes *symétrique*.

Notons que l'on a

$$\hat{a} \sim \frac{1}{n} \sum_{k=1}^{n} x_k, \tag{2.169}$$

pour

$$\sigma_0^2 \gg \frac{\sigma^2}{n}. \tag{2.169'}$$

La condition (2.169′) se trouve remplie si pour $\left(\frac{\sigma}{\sigma_0}\right)^2$ donné, la taille n de l'échantillon augmente indéfiniment ou si pour n donné on a $\sigma_0^2 \gg \sigma^2$. La première condition signifie que l'estimation de Bayes (2.167) pour $n \to \infty$ tend asymptotiquement vers l'estimation du maximum de vraisemblance. La seconde condition peut être interprétée comme suit : la densité de probabilité $w_1(a)$ du paramètre inconnu est à peu près uniforme lorsqu'on la compare à la densité de probabilité initiale $w_1(x; a)$. Par conséquent, l'estimation (2.167) pour $\sigma_0^2 \gg \sigma^2$ devient l'estimation du maximum de vraisemblance, ce qui a déjà été noté sous une forme générale à la fin du § 2.3.3.

Notons que dans le cas inverse, lorsque la variance σ^2 de la distribution initiale σ^2 est très supérieure à $n\sigma_0^2$, on obtient à partir de (2.167) la relation asymptotique suivante

$$\hat{a} \sim a_0,$$

ce qui veut dire que les valeurs échantillonnées n'influent pas sur l'estimation, qui est prise égale à la moyenne du paramètre estimé.

Calculons maintenant la valeur du risque moyen pour une fonction de pertes quadratique. En portant (2.167) et (2.168) dans (2.79) on obtient

$$J(\hat{a} \mid x_1, \ldots, x_n) = \sqrt{\frac{n}{2\pi\sigma^2}} \left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)^{1/2} \times$$

$$\times \int_{-\infty}^{\infty} (a - \hat{a})^2 \exp\left\{-\frac{n}{2\sigma^2}\left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)(a - \hat{a})^2\right\} da =$$

$$= \frac{\sigma^2}{n}\left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)^{-1},$$

ce qui signifie que la fonctionnelle en question est une grandeur constante. Puis on tire de (2.78)

$$R = \frac{\sigma^2}{n}\left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)^{-1}. \tag{2.170}$$

Le maximum du risque moyen correspond au cas où $\sigma_0 \to \infty$, par) conséquent la *distribution uniforme* (dans le sens mentionné ci-dessus

du paramètre a *est la moins favorable* et

$$R_{\max} = \frac{\sigma^2}{n}. \qquad (2.171)$$

Pour cette distribution, on peut trouver la fonction de risque en portant (2.169) dans (2.76) et compte tenu de l'indépendance des variables x_i et x_k pour $i \neq k$, on a

$$r(a) = \frac{1}{n^2} \sum_{k=1}^{n} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (x_k - a)^2 L_x(a)\, dx_1 \ldots dx_n = \frac{\sigma^2}{n}. \qquad (2.172)$$

En vertu de (2.172) la fonction de risque ne dépend pas du paramètre estimé et sa valeur coïncide avec le maximum du risque moyen.

Ainsi, la moyenne échantillonnée est en même temps *l'estimation du minimax* de la moyenne d'une variable aléatoire normale lorsque l'on connaît exactement sa variance et que l'on sait que la fonction de pertes est quadratique.

Soit maintenant l'estimation de Bayes de la variance σ^2 de la loi normale, la moyenne a étant donnée (posons $a = 0$). Supposons que la densité de probabilité a priori de la variance soit exponentielle de paramètre σ_0^2

$$w_1(\sigma^2) = \frac{1}{\sigma_0^2} e^{-\frac{\sigma^2}{\sigma_0^2}}. \qquad (2.173)$$

Nous allons trouver l'estimation correspondant au maximum de densité de probabilité a posteriori du paramètre estimé σ^2. En utilisant (2.69) on obtient

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} x_i^2 - \frac{1}{\sigma_0^2} = 0,$$

d'où

$$\hat{\sigma}^2 = \frac{n\sigma_0^2}{4} \left[\sqrt{1 + \frac{8}{n^2} \sum_{i=1}^{n} \left(\frac{x_i}{\sigma_0} \right)^2} - 1 \right]. \qquad (2.174)$$

Pour $n\sigma_0^2 \to \infty$ on a en vertu de (2.174)

$$\hat{\sigma}^2 \sim \frac{1}{n} \sum_{i=1}^{n} x_i^2. \qquad (2.175)$$

Ainsi, lorsque la taille de l'échantillon augmente indéfiniment ou encore pour n donné, au fur et à mesure que la fonction (2.173) tend vers la distribution uniforme, l'estimation de la variance correspondant au maximum de densité a posteriori tend asymptotiquement vers la variance échantillonnée, c'est-à-dire vers l'estimation du maximum de vraisemblance.

Pour une fonction de pertes quadratique l'estimation de Bayes est [cf. (2.87)]

$$\hat{\sigma}^2 = m_1\{\sigma^2 \mid x_1, \ldots, x_n\} =$$

$$= \frac{\int\limits_0^\infty z^{-\frac{n}{2}+1} \exp\left\{-\left(\frac{z}{\sigma_0^2}+\frac{1}{2z}\sum_{i=1}^n x_i^2\right)\right\} dz}{\int\limits_0^\infty z^{-\frac{n}{2}} \exp\left\{-\left(\frac{z}{\sigma_0^2}+\frac{1}{2z}\sum_{i=1}^n x_i^2\right)\right\} dz}. \quad (2.176)$$

En utilisant (2.136) on peut écrire (2.176) comme suit:

$$\hat{\sigma}^2 = \sigma_0^2 \sqrt{\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i}{\sigma_0}\right)^2} \times \frac{K_{\frac{n}{2}-2}\left[\sqrt{2\sum_{i=1}^n \left(\frac{x_i}{\sigma_0}\right)^2}\right]}{K_{\frac{n}{2}-1}\left[\sqrt{2\sum_{i=1}^n \left(\frac{x_i}{\sigma_0}\right)^2}\right]}. \quad (2.177)$$

Comme dans le cas de l'estimation du paramètre d'une loi exponentielle, l'asymétrie de la densité de probabilité a posteriori de la variance fait que l'estimation de Bayes (2.177) pour une fonction de pertes quadratique diffère de l'estimation de Bayes (2.174) correspondant à une fonction de pertes simple.

2.6. ESTIMATION D'UNE DISTRIBUTION UNIDIMENSIONNELLE

2.6.1. Critère de conformité. Dans les paragraphes précédents nous avons étudié les méthodes d'estimation des paramètres inconnus de la distribution d'une variable aléatoire, supposant connue la loi de distribution. Souvent, a priori, on ne sait absolument rien sur la forme de la loi en question et il faut, connaissant des valeurs échantillonnées, trouver la forme de la distribution d'où l'échantillon a été extrait.

En principe, le théorème de Glivenko (cf. § 2.1.1) permet de résoudre le problème. En vertu de ce théorème la fonction de répartition empirique $F_1^*(x)$ converge en probabilité vers la fonction hypothétique $F_1(x)$ si l'échantillon ayant servi à trouver la fonction empirique est issu de $F_1(x)$. Mais ce théorème, établissant la tendance asymptotique de $F_1^*(x)$ vers $F_1(x)$ pour $n \to \infty$, ne permet pas, pour un échantillon de *taille n finie*, de trouver les probabilités d'écarts entre la répartition empirique et la répartition hypothétique. Dans ce cas, il y a lieu avant tout d'introduire une *mesure quantitative de conformité* de la répartition hypothétique et de la répartition empirique ou, selon la terminologie usitée, le *critère de confor-*

mité. Ce critère est le nombre $\Delta(F_1^*, F_1)$ qui est une fonctionnelle des répartitions empirique $F_1^*(x)$ et hypothétique $F_1(x)$.

Ce problème peut être résolu de la manière suivante. A partir des valeurs échantillonnées $x_1, \ldots, x_n$ on trouve à l'aide de la formule (2.1) la fonction de répartition empirique $F_1^*(x)$. A partir de cette répartition et peut-être à la base de certaines autres considérations, on émet l'hypothèse que l'échantillon provient de la répartition $F_1(x)$. Selon le critère de conformité adopté à l'avance, il y a lieu de calculer $\Delta(F_1^*, F_1)$ ainsi que la probabilité pour cette grandeur de dépasser un certain seuil Δ_0. A cet effet, il faut connaître la répartition $\Delta(F_1^*, F_1)$. En général, on arrive à trouver cette répartition sous une forme assez simple seulement pour des échantillons de taille suffisamment grande (plus exactement on arrive à trouver la répartition asymptotique pour $n \to \infty$). Connaissant la répartition de la grandeur $\Delta(F_1^*, F_1)$, on peut, s'étant donné la probabilité α (niveau de signification du critère) d'avoir $\Delta > \Delta_0$, trouver le seuil Δ_0. Pour α suffisamment petit, on obtient une règle assez bonne de vérification de l'hypothèse sur la répartition $F_1(x)$: si pour l'échantillon extrait on a $\Delta > \Delta_0$, l'hypothèse est rejetée, dans le cas contraire elle est adoptée. La probabilité de rejeter une hypothèse vraie sera alors égale au niveau de signification (voir chapitre 1, page 24).

Notons que la répartition de la grandeur $\Delta(F_1^*, F_1)$ déterminée à partir de la condition que l'échantillon $x_1, \ldots, x_n$ ayant servi à trouver $F_1^*(x)$ a été extrait de $F_1(x)$, dépend en général de $F_1(x)$. Il serait souhaitable d'avoir des critères de conformité dont la répartition ne dépendrait pas de la forme de la fonction hypothétique $F_1(x)$. De tels critères sont appelés *non paramétriques* (cf. § 1.5.3). Les cas particuliers des critères de conformité étudiés ci-dessous sont asymptotiquement (pour $n \to \infty$) non paramétriques.

Il ne faut pas oublier non plus que le critère de conformité, choisi à partir de certaines considérations, permet de formuler, pour un niveau de signification α suffisamment bas, une règle acceptable de *rejet d'une hypothèse fausse*. L'hypothèse vraie se trouvera rejetée avec une fréquence égale à α. Cependant, dans ce cas la fréquence des erreurs dues à l'adoption d'une hypothèse fausse reste indéterminée. Si pour l'échantillon donné on a $\Delta < \Delta_0$, bien que la règle adoptée impose d'accepter l'hypothèse avancée, il y a lieu de faire attention. En effet, la condition $\Delta < \Delta_0$ veut dire simplement qu'au point de vue du critère de conformité adopté, il n'y a aucune raison de considérer que les valeurs échantillonnées ne soient pas conformes à la répartition hypothétique.

L'inconvénient des critères de conformité est que le critère lui-même et le niveau de signification sont arbitraires, de plus il est assez difficile de trouver leurs répartitions pour des échantillons de taille finie. Une autre méthode consiste à trouver directement les estimations de la fonction de répartition inconnue d'après des va-

leurs échantillonnées, sans émettre d'hypothèse concernant la forme de cette fonction (fonction de répartition empirique uniformisée).

2.6.2. Critère du khi carré. Ce critère de conformité est l'un des plus répandus dans les applications.

Divisons le domaine de définition de la fonction de répartition hypothétique $F_1(x)$ de la variable aléatoire ξ en un nombre fini d'intervalles disjoints Δ_i, $i = 1, \ldots, l$. Introduisons la notation

$$p_i = P\{\xi \in \Delta_i\}, \quad \sum_{i=1}^{l} p_i = 1. \qquad (2.178)$$

Supposons que dans l'échantillon $x_1, \ldots, x_n$, le nombre d'éléments tombés dans l'intervalle Δ_i soit égal à ν_i. Il est évident que l'on a $\sum_{i=1}^{l} \nu_i = n$.

Prenons pour le critère de conformité la grandeur suivante:

$$\Delta = \sum_{i=1}^{l} \frac{n}{p_i}\left(\frac{\nu_i}{n} - p_i\right)^2 = \sum_{i=1}^{l} \frac{1}{np_i}(\nu_i - np_i)^2. \qquad (2.179)$$

Ainsi que l'a montré Pearson (cf. [5], § 30.1), si l'hypothèse sur l'authenticité de la répartition $F_1(x)$ est vraie, pour $n \to \infty$ la répartition du critère (2.179) tend asymptotiquement vers la loi du χ^2 à $l - 1$ degrés de liberté et *ne dépend pas de la forme de la répartition hypothétique* $F_1(x)$.

Soit χ^2_α le quantile de la variable aléatoire répartie suivant la loi du χ^2 à $l - 1$ degrés de liberté, c'est-à-dire $P\{\chi^2 > \chi^2_\alpha\} = \alpha$ (voir annexe VIII). Lorsque α est suffisamment petit et la taille de l'échantillon est suffisamment grande, la grandeur Δ calculée à l'aide de (2.179) ne surpasse pratiquement jamais le seuil χ^2_α, à condition toutefois que l'hypothèse concernant la forme de la répartition de laquelle l'échantillon a été extrait soit vraie. On adopte ainsi la règle suivante de vérification de l'hypothèse: l'hypothèse est rejetée si $\Delta > \chi^2_\alpha$ et adoptée si $\Delta \leqslant \chi^2_\alpha$. La probabilité de rejeter une hypothèse vraie est égale à α.

En plus des défauts généraux mentionnés plus haut, inhérents à tous les critères de conformité, le critère du χ^2 présente l'inconvénient d'une partition arbitraire du domaine des valeurs possibles de la variable aléatoire en intervalles Δ_i, n'étant pas dictée ni par la forme de la fonction $F_1(x)$, ni par un groupement adéquat des éléments.

2.6.3. Critère de Kolmogorov. En vertu de ce critère, la mesure quantitative de conformité, pour un échantillon de taille n donnée, est le maximum pour toutes les valeurs x du module de l'écart de la répartition empirique par rapport à la répartition hypothétique,

c'est-à-dire

$$\Delta = \max_x | F_1^* (x) - F_1 (x) |. \tag{2.180}$$

A. Kolmogorov a montré que (cf. [4], § 62) si l'hypothèse sur l'authenticité de $F_1 (x)$ se trouve vérifiée, pour $n \to \infty$ et, de plus, pour $F_1 (x)$ continue, la fonction de répartition de la grandeur $\Delta\sqrt{n}$ tend asymptotiquement vers *)

$$P\{\Delta \sqrt{n} \leqslant z\} \sim \mathrm{k}(z) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2z^2}, \quad z > 0. \tag{2.181}$$

La fonction k (z) peut s'écrire comme suit:

$$\mathrm{k}(z) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2z^2}, \quad z > 0. \tag{2.182}$$

D'un autre côté, en écrivant k (z) sous la forme

$$\mathrm{k}(z) = \sum_{k=-\infty}^{\infty} e^{-2(2k)^2z^2} - \sum_{k=-\infty}^{\infty} e^{-2(2k+1)^2z^2}$$

et en utilisant la formule de sommation de Poisson

$$\sum_{k=-\infty}^{\infty} h(k) = \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) e^{2\pi ikx} dx,$$

on peut écrire cette fonction comme suit:

$$\mathrm{k}(z) = \frac{\sqrt{2\pi}}{z} \sum_{k=1}^{\infty} e^{-\frac{\pi^2(2k-1)^2}{8z^2}}. \tag{2.183}$$

Il est commode d'utiliser la formule (2.182) pour $z > 1$, et la formule (2.183) pour $z \leqslant 1$.

Soit α un niveau de signification donné, il vient de (2.182)

$$P\{\Delta > \Delta_\alpha\} = P\{\Delta \sqrt{n} > \Delta_\alpha \sqrt{n}\} \sim 1 - \mathrm{k}(\Delta_\alpha \sqrt{n}) =$$

$$= 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2n\Delta_\alpha^2} = \alpha \tag{2.184}$$

ou sous une autre forme, compte tenu de (2.183),

$$P\{\Delta > \Delta_\alpha\} \sim \frac{\sqrt{2\pi}}{\Delta_\alpha \sqrt{n}} \sum_{k=1}^{\infty} e^{-\frac{\pi^2(2k-1)^2}{8n\Delta_\alpha^2}} = \alpha. \tag{2.185}$$

*) La fonction de répartition envisagée peut s'écrire comme suit [12]:

$$P\{\Delta \sqrt{n} \leqslant z\} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2z^2} \left[1 - \frac{2k^2x}{3\sqrt{n}} + o\left(\frac{1}{n}\right)\right].$$

La série dans (2.184) étant rapidement convergente, pour première approximation on peut souvent se limiter au premier terme, c'est-à-dire

$$2e^{-2n\Delta_\alpha^2} = \alpha \tag{2.186}$$

ou

$$\Delta_\alpha = \sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}. \tag{2.186'}$$

On obtient la règle suivante de vérification de l'hypothèse: si pour l'échantillon observé on a $\Delta > \Delta_\alpha$, l'hypothèse selon laquelle l'échantillon provient de la répartition hypothétique est rejetée, dans le cas contraire on adopte cette hypothèse. La table des valeurs de la fonction k (z) est donnée dans [4]. Tout comme le critère du χ^2, le critère de Kolmogorov est utilisé lorsque l'échantillon est de taille importante. Cependant, pour ce critère on n'a pas besoin de procéder à une partition préalable de l'intervalle des valeurs ni à un groupement des éléments, comme c'était nécessaire dans le cas du critère du χ^2.

Notons que pour un niveau de signification donné, la fonction de répartition hypothétique se tient à l'intérieur d'une bande dont les limites sont données par les équations

$$y = F_1^*(x) + \Delta$$

et

$$y = F_1^*(x) - \Delta.$$

2.6.4. Critère de Mises. En vertu de ce critère, la mesure quantitative de conformité, pour un échantillon de taille n donnée, est la valeur moyenne du carré de l'écart de la répartition empirique par rapport à la répartition hypothétique, soit:

$$\Delta = \int_{-\infty}^{\infty} [F_1^*(x) - F_1(x)]^2 w_1(x)\, dx, \tag{2.187}$$

ici $w_1(x) = F_1'(x)$ est la densité de probabilité hypothétique de la variable aléatoire. En portant (2.1) dans (2.187) et en intégrant on obtient la grandeur Δ sous la forme suivante *):

$$\Delta = \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^{n} \left[F_1(x_k) - \frac{2k-1}{2n}\right]^2. \tag{2.187'}$$

Il est facile de voir [3] que

$$m_1\{\Delta\} = \frac{1}{6n}, \quad M_2\{\Delta\} = \frac{4n-3}{180n^3}. \tag{2.188}$$

*) Souvent dans ce cas, au lieu de Δ on utilise la désignation ω^2 et on appelle le critière de conformité critère ω^2.

L'expression de la répartition exacte de la grandeur $n\Delta$ est très compliquée, mais pour $n > 40$ elle est voisine d'une certaine répartition limite. A la différence du critère du χ^2, le critère de Mises, comme le critère de Kolmogorov, n'exige pas de groupement des éléments de l'échantillon et sa répartition se rapproche assez vite, avec augmentation de la taille de l'échantillon, d'une répartition limite. Ci-dessous on trouve quelques quantiles de cette répartition limite:

γ	0,5	0,4	0,3	0,2	0,1	0,05
z_γ	0,1184	0,1467	0,1843	0,2412	0,3473	0,4614
γ	0,03	0,02	0,01	0,001		
z_γ	0,5489	0,6198	0,7435	1,1679		

2.6.5. Appartenance de deux échantillons à une même répartition. Supposons que l'on ait deux échantillons $(x_1, \ldots, x_n)$ et $(y_1, \ldots \ldots, y_m)$ à éléments indépendants, appartenant chacun à une répartition quelconque. Il est intéressant de vérifier l'hypothèse selon laquelle les deux échantillons appartiennent à *une même* répartition. Soient $F_{1x}^*(z)$ et $F_{1y}^*(z)$ les fonctions de répartition empiriques construites d'après ces échantillons. On prend la grandeur

$$\Delta = \max_z |F_{1x}^*(z) - F_{1y}^*(z)| \qquad (2.189)$$

en qualité de critère de conformité. Comme l'a montré N. Smirnov (cf. [4], § 63), pour $n \to \infty$ on a

$$P\left\{\Delta\left(\frac{1}{n}+\frac{1}{m}\right)^{-\frac{1}{2}} > z\right\} \sim 1 - \mathrm{k}(z), \qquad (2.190)$$

où $\mathrm{k}(z)$ est une fonction donnée par (2.182).

Si α est un niveau de signification donné, d'une manière analogue à (2.186′) on peut en première approximation calculer le seuil

$$\Delta_\alpha = \sqrt{\frac{1}{n}+\frac{1}{m}} \cdot \sqrt{\frac{1}{2}\ln\frac{2}{\alpha}} \qquad (2.190')$$

et formuler la règle suivante de vérification de l'hypothèse: les échantillons appartiennent à une même répartition si $\Delta \leqslant \Delta_\alpha$.

Un autre critère, proposé par Wilcoxon, tient compte du nombre d'*inversions*. A cet effet on compose à partir des deux échantillons un échantillon *ordonné* (voir § 2.1.1)

$$y_1,\ y_2,\ x_1,\ y_3,\ x_2,\ \ldots$$

Si dans cette suite à un x_i donné précèdent s éléments $y_1, \ldots, y_s$, de l'autre échantillon on a s inversions. Le nombre total d'inversions U est égal à la somme des inversions formées par tous les éléments du premier échantillon avec les éléments du second. La règle de vérification d'une hypothèse suivant le critère de Wilcoxon consiste à comparer le nombre total d'inversions avec le seuil déterminé par le niveau de signification donné.

On peut montrer (cf. [2], page 357) que pour $m+n \geqslant 20$ et $m > 3$ on peut considérer, sans commettre de grande erreur, que le nombre total d'inversions est une variable *normale* de paramètres

$$m_1\{U\} = \frac{mn}{2}, \quad M_2\{U\} = \frac{mn}{12}(m+n+1). \qquad (2.191)$$

La valeur de seuil U_α du nombre d'inversions, pour un niveau de signification α donné, se trouve comme suit

$$U_\alpha = \frac{mn}{2} + x_\alpha \sqrt{\frac{mn}{12}(m+n+1)}, \qquad (2.192)$$

où x_α est le quantile d'une variable aléatoire normale. Si la valeur de U calculée d'après les deux échantillons donnés est supérieure à U_α, on rejette l'hypothèse selon laquelle ces échantillons appartiennent à une même répartition.

On peut également utiliser la variante à deux seuils du critère de Wilcoxon où le domaine critique des valeurs de U dans lequel on rejette l'hypothèse formulée ci-dessus est donné par les deux seuils :

$$U_{\alpha\,\inf} = \frac{mn}{12} - x_{\frac{\alpha}{2}} \sqrt{\frac{mn}{12}(m+n+1)}, \qquad (2.193)$$

$$U_{\alpha\,\sup} = \frac{mn}{12} + x_{\frac{\alpha}{2}} \sqrt{\frac{mn}{12}(m+n+1)}. \qquad (2.193')$$

2.6.6. Estimation d'une densité de probabilité. Soit $x_1, \ldots, x_n$ un échantillon à éléments indépendants, extrait d'une distribution inconnue de densité de probabilité $w_1(x)$. Pour estimer cette densité on peut utiliser la fonction suivante (fig. 2.8) :

$$\hat{w}_1(x) = \frac{1}{nh(n)} \sum_{i=1}^{n} K\left[\frac{x-x_i}{h(n)}\right], \qquad (2.194)$$

où $K(y)$ est un noyau quelconque d'approximation satisfaisant aux conditions

$$0 < K(y) < \infty, \quad \lim_{y\to\pm\infty} yK(y) = 0, \quad \int_{-\infty}^{\infty} K(y)\,dy = 1, \qquad (2.194')$$

et

$$\lim_{n\to\infty} h(n) = 0. \qquad (2.194'')$$

Il a été démontré (cf. [15], [17]) que l'estimation (2.194) est consistante et *asymptotiquement non biaisée.* Le choix de la forme du noyau $K(y)$ peut être astreint à certaines conditions supplémentaires d'optimalité de l'estimation. Dans certains cas il est bon de prendre un noyau gaussien

$$K(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}.$$

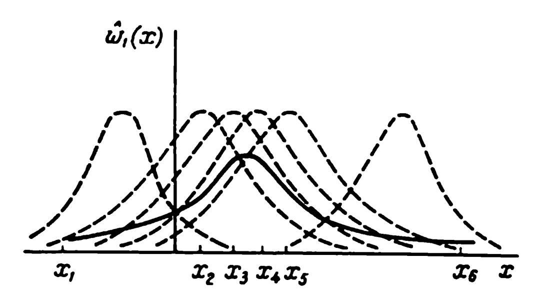


Fig. 2.8. Construction de l'estimation d'une fonction densité de probabilité

Si l'on connaît la forme de la fonction $w_1(x)$, mais on ne connaît pas ses paramètres, pour l'estimation de cette densité de probabilité on peut prendre la fonction $\hat{w}_1(x)$ dont les paramètres sont des estimations des paramètres inconnus obtenues à partir des valeurs échantillonnées. Par exemple, pour estimer une densité normale de paramètres inconnus, on peut utiliser la fonction

$$\hat{w}_1(x)=\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}e^{-\frac{(x-\hat{a})^2}{2\hat{\sigma}^2}} \qquad (2.195)$$

où $\hat{a}$, $\hat{\sigma}^2$ sont les estimations déterminées respectivement par (2.139) et (2.140).

2.7. ESTIMATION DES PARAMÈTRES DES DISTRIBUTIONS MULTIDIMENSIONNELLES

2.7.1. Généralisation des définitions fondamentales au cas des distributions multidimensionnelles. Tout ce qui a été exposé ci-dessus concerne la théorie des estimations d'une seule variable aléatoire. Les estimations des distributions multidimensionnelles ou de leurs paramètres pour un ensemble de variables aléatoires *dépendantes* $\xi_1, \ldots, \xi_N$ présentent également un grand intérêt pratique. Dans la théorie paramétrique on connaît la forme de la fonction de répartition multidimensionnelle de cet ensemble et l'on estime seulement les paramètres inconnus, tandis que dans la théorie non paramétrique c'est la fonction de répartition elle-même qui est estimée. Dans les deux cas les déductions statistiques sont faites à partir d'un *échantillon* provenant d'une distribution *multidimensionnelle.* L'élément d'un tel échantillon (de taille n) n'est pas un nombre comme dans le cas unidimensionnel, mais N nombres, soit

$$x_{1k},\ x_{2k},\ \ldots,\ x_{Nk},\ k=1,\ \ldots,\ n.$$

Autrement dit, on peut représenter un élément d'un échantillon d'une distribution multidimensionnelle par un vecteur $\mathbf{x}_k$ ayant pour com-

posantes les nombres ci-dessus, et tout l'échantillon, par une matrice rectangulaire $N \times n$

$$\mathbf{X} = \| x_{ik} \|.$$

La fonction de vraisemblance de l'échantillon d'une distribution à N dimensions est une fonction de n arguments vectoriels, les paramètres de cette fonction étant des matrices (des vecteurs en particulier) dont les éléments sont les paramètres inconnus de la distribution multidimensionnelle d'un ensemble de variables aléatoires. Pour les éléments indépendants d'un échantillon multidimensionnel on a

$$L_X(\vartheta) = \prod_{k=1}^{n} W_N(x_{1k}, \ldots, x_{Nk}; \vartheta_1, \ldots, \vartheta_m), \qquad (2.196)$$

où W_N est la densité de probabilité de l'ensemble des variables aléatoires $\xi_1, \ldots, \xi_N$.

A partir de l'échantillon $\mathbf{X}$ on trouve s matrices échantillonnées (ou s vecteurs) dépendant des vecteurs échantillonnés $\mathbf{x}_1, \ldots, \mathbf{x}_n$:

$$\hat{\mathbf{M}}_i = g(\mathbf{x}_1, \ldots, \mathbf{x}_n), \quad i = 1, \ldots, s, \qquad (2.197)$$

en supposant que les paramètres de la densité W_N soient donnés. Ces matrices (vecteurs) échantillonnées sont des *estimations conditionnelles* des matrices $\mathbf{M}_i$ (vecteurs) dont les éléments sont les paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$.

Si les paramètres de la distribution sont aléatoires et si l'on connaît leur distribution conjointe a priori, on peut calculer des *estimations inconditionnelles*.

Chacune des estimations conditionnelles $\hat{\mathbf{M}}_i$ [cf. (2.197)] est dite consistante si pour $n \to \infty$ elle converge en probabilité vers $\mathbf{M}_i$. L'estimation $\hat{\mathbf{M}}_i$ est dite non biaisée si pour n quelconque sa valeur moyenne sur l'ensemble des échantillons vectoriels est égale à $\mathbf{M}_i$. Il ne faut pas oublier que l'on entend par valeur moyenne d'une matrice (d'un vecteur) aléatoire la matrice (le vecteur) dont les éléments sont égaux aux valeurs moyennes des éléments d'une matrice (d'un vecteur) aléatoire. D'une manière analogue on peut généraliser la notion d'estimations simultanément suffisantes. A cet effet les arguments scalaires dans (2.97) sont remplacés par des arguments vectoriels. La matrice informationnelle de Fisher $\mathbf{I}_n$ est une matrice partitionnée dont les éléments sont les matrices

$$\mathbf{I}_n^{(i,j)}(\mathbf{M}_1, \ldots, \mathbf{M}_s) = m_1 \left\{ \frac{\partial}{\partial \mathbf{M}_i} \ln L_X(\mathbf{M}) \frac{\partial}{\partial \mathbf{M}_j} \ln L_X(\mathbf{M}) \right\}, \qquad (2.198)$$

$$i, j = 1, \ldots, s.$$

Les estimations au maximum de vraisemblance peuvent être obtenues à partir du système d'équations

$$\frac{\partial \ln L_X(\mathbf{M})}{\partial \mathbf{M}_i} = 0, \quad i = 1, \ldots, s. \qquad (2.199)$$

De même les estimations correspondant au maximum de densité de probabilité a posteriori des paramètres estimés sont données par

$$\frac{\partial \ln w_s(\mathbf{M}_1, \ldots, \mathbf{M}_s)}{\partial \mathbf{M}_l} + \frac{\partial \ln L_{\mathbf{X}}(\mathbf{M})}{\partial \mathbf{M}_l} = 0. \qquad (2.199')$$

2.7.2. Estimations du vecteur des valeurs moyennes et de la matrice de corrélation d'une distribution normale multidimensionnelle. Avant d'introduire les estimations de Bayes des paramètres d'une distribution multidimensionnelle nous allons illustrer les généralisations mentionnées ci-dessus sur l'exemple d'une distribution normale à N dimensions. Ecrivons la densité de probabilité de cette distribution sous la forme vectorielle (cf. (2.57), tome I):

$$W_N(\mathbf{X}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \mathbf{M}}} \exp\left[-\frac{1}{2}(\mathbf{x}' - \mathbf{a}')\mathbf{M}^{-1}(\mathbf{x} - \mathbf{a})\right] \qquad (2.200)$$

où $\mathbf{a}$ est le vecteur des valeurs moyennes, $\mathbf{M}$ la matrice de corrélation.

Soient $\mathbf{x}_1, \ldots, \mathbf{x}_n$ des vecteurs échantillonnés indépendants à N dimensions issus de la distribution normale considérée. La fonction de vraisemblance (2.196) de l'échantillon est égale à

$$L_{\mathbf{X}}(\mathbf{a}, \mathbf{M}) = \prod_{k=1}^{n} W_N(\mathbf{x}_k;\ \mathbf{a}, \mathbf{M}) = \frac{1}{(2\pi)^{\frac{nN}{2}} (\det \mathbf{M})^{\frac{n}{2}}} \exp\left[-\frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \mathbf{a})'\mathbf{M}^{-1}(\mathbf{x}_k - \mathbf{a})\right]. \qquad (2.201)$$

Introduisons le vecteur des moyennes échantillonnées

$$\mathbf{m}_1^* = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k \qquad (2.202)$$

et la matrice de corrélation échantillonnée

$$\mathbf{M}^* = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \mathbf{m}_1^*)(\mathbf{x}_k - \mathbf{m}_1^*)'. \qquad (2.203)$$

On peut montrer que (cf., par exemple, [1], § 3.2)

$$\sum_{k=1}^{n}(\mathbf{x}_k - \mathbf{a})'\mathbf{M}^{-1}(\mathbf{x}_k - \mathbf{a}) = \operatorname{tr}(n\mathbf{M}^{-1}\mathbf{M}^*) + n(\mathbf{m}_1^* - \mathbf{a})'\mathbf{M}^{-1}(\mathbf{m}_1^* - \mathbf{a}). \qquad (2.204)$$

où le symbole tr désigne la trace de la matrice *).

*) On appelle trace de la matrice carrée $A = \| a_{ij} \|$ la somme de ses éléments diagonaux: $\operatorname{tr} A = \sum_i a_{ii}$. Parfois on utilise la notation sp A.

En portant (2.204) dans (2.201) on peut écrire comme suit la fonction de vraisemblance:

$$L_{\mathbf{X}}(\mathbf{a}, \mathbf{M}) = \frac{1}{(2\pi)^{\frac{nN}{2}} (\det \mathbf{M})^{\frac{n}{2}}} \exp\left[-\frac{1}{2} n (\mathbf{m}_1^* - \mathbf{a})' \mathbf{M}^{-1} (\mathbf{m}_1^* - \mathbf{a})\right] \times$$
$$\times \exp\left[-\frac{1}{2} \operatorname{tr}(n\mathbf{M}^{-1}\mathbf{M}^*)\right]. \qquad (2.205)$$

En vertu de (2.205) $\mathbf{m}_1^*$ et $\mathbf{M}^*$ sont des estimations simultanément suffisantes du vecteur des valeurs moyennes et de la matrice de corrélation d'une distribution normale multidimensionnelle. Ces mêmes estimations sont les estimations du maximum de vraisemblance. L'estimation $\hat{\mathbf{a}} = \mathbf{m}_1^*$ est absolument correcte et efficace. L'estimation $\mathbf{M}^*$ est biaisée, en effet

$$m_1\{\mathbf{M}^*\} = \left(1 - \frac{1}{n}\right) \mathbf{M}. \qquad (2.206)$$

C'est pourquoi pour l'estimation absolument correcte de la matrice de corrélation $\mathbf{M}$ on a

$$\hat{\mathbf{M}} = \frac{n}{n-1} \mathbf{M}^* = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{x}_k - \mathbf{m}_1^*)(\mathbf{x}_k - \mathbf{m}_1^*)'. \qquad (2.207)$$

L'efficacité des estimations $\hat{\mathbf{a}}$ et $\hat{\mathbf{M}}$ (c'est-à-dire le rapport des volumes de l'ellipsoïde informationnel et de l'ellipsoïde de corrélation) est égale à $\left(\frac{n-1}{n}\right)^{\frac{N(N+1)}{2}}$.

2.7.3. Estimations de Bayes des paramètres d'une distribution normale multidimensionnelle. Nous allons généraliser (2.109). A cet effet introduisons la fonction de pertes pour l'estimation des matrices d'une distribution multidimensionnelle, soit:

$$\Pi = \Pi(\hat{\mathbf{M}}_1, \ldots, \hat{\mathbf{M}}_s, \mathbf{M}_1, \ldots, \mathbf{M}_s). \qquad (2.208)$$

Les estimations conditionnelles de Bayes minimisent la fonction de risque conditionnelle

$$r(\mathbf{M}_1, \ldots, \mathbf{M}_s) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Pi(\hat{\mathbf{M}}_1, \ldots, \hat{\mathbf{M}}_s, \mathbf{M}_1, \ldots, \mathbf{M}_s) \times$$
$$\times L_{\mathbf{X}}(\mathbf{M})\, dx_1 \ldots dx_n, \qquad (2.209)$$

Dans le cas où les paramètres estimés $\mathbf{M}_1, \ldots, \mathbf{M}_s$ sont aléatoires, en introduisant leur densité de probabilité a priori $w_s(\mathbf{M}_1, \ldots, \mathbf{M}_s)$ on peut écrire comme suit la valeur du risque moyen:

$$R = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} w_s(\mathbf{M}_1, \ldots, \mathbf{M}_s)\, r(\mathbf{M}_1, \ldots, \mathbf{M}_s)\, d\mathbf{M}_1 \ldots d\mathbf{M}_s. \quad (2.210)$$

Les estimations inconditionnelles de Bayes peuvent être obtenues en minimisant la fonctionnelle matricielle (vectorielle) (2.210).

2.7.4. Remarque concernant les estimations d'une distribution multidimensionnelle. Dans le cas des distributions multidimensionnelles les critères de conformité n'ont pas fait l'objet d'études spéciales.

La formule (2.194) peut être étendue tout naturellement au cas multidimensionnel pour estimer une densité de probabilité inconnue $W_N(x_1, \ldots, x_N)$.

Soient $\mathbf{x}_1, \ldots, \mathbf{x}_n$ les vecteurs des valeurs échantillonnées appartenant à la distribution mentionnée, et $x_{1k}, \ldots, x_{Nk}$ les composantes du k-ième vecteur. Si pour $k \neq l$, x_k et x_l sont indépendants, l'estimation de W_N peut être donnée par la fonction

$$\hat{W}_N(x_1, \ldots, x_N) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{N} \frac{1}{h_j(n)} K_j\left[\frac{x_j - x_{jl}}{h_j(n)}\right], \qquad (2.211)$$

où $K_j(y)$ est le noyau de l'approximation satisfaisant aux conditions suivantes ($j = 1, \ldots, N$)

$$0 < K_j(y) < \infty, \quad \lim_{y \to \pm\infty} yK_j(y) = 0,$$

$$\int_{-\infty}^{\infty} K_j(y)\,dy = 1, \quad \lim_{n \to \infty} h_j(n) = 0. \qquad (2.211')$$

L'estimation (2.211) est consistante et asymptotiquement correcte.

Problèmes

2.1. Montrer que la moyenne et la variance échantillonées d'une variable aléatoire normale sont indépendantes. Généraliser ce résultat au cas d'une distribution normale multidimensionnelle et montrer que le vecteur des moyennes échantillonnées ne dépend pas de la matrice des variances-covariances échantillonnée.

2.2. Montrer que pour un coefficient de confiance γ donné, l'intervalle de confiance minimal pour une variance inconnue d'une variable aléatoire normale doit satisfaire à la condition (fig. 2.9)

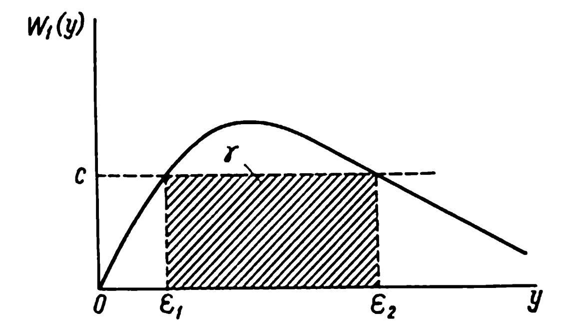


Fig. 2.9. Détermination de l'intervalle de confiance minimal

$$W_1(y) \geqslant c, \tag{1}$$

où $W_1(y)$ est la densité de probabilité de la loi du χ^2 à $n-1$ degrés de liberté [cf. (2.157)], et la constante c est donnée par la condition

$$\int_{\varepsilon_1}^{\varepsilon_2} W_1(y)\,dy = \gamma, \tag{2}$$

où ε_1 et ε_2 sont les projections sur l'axe des abscisses des points d'intersection de la courbe $z = W_1(y)$ et de la droite $z = c$ [cf. (2.158)].

2.3. Supposons que l'on effectue k groupes d'observations indépendantes, dont le résultat est k échantillons de tailles $n_1, \ldots, n_k$ prélevés sur des répartitions continues $F_{1i}(x)$, $i = 1, \ldots, k$ (cf. [14]). On vérifie l'hypothèse selon laquelle $F_{1i}(x) = F_1(x)$ pour tous les i, ayant choisi pour critère de conformité la grandeur

$$\Delta = \max_{-\infty<x<\infty} \Big[\sum_{i=1}^{k} n_i \{F_{1i}^*(x) - F_0^*(x)\}^2\Big]^{\frac{1}{2}}, \tag{3}$$

où

$$F_0^* = \frac{\sum\limits_{i=1}^{k} n_i F_{1i}^*(x)}{\sum\limits_{i=1}^{k} n_i}; \tag{4}$$

$F_{1i}^*(x)$ est la fonction de répartition empirique construite d'après l'i-ième échantillon. Montrer que pour $n_1, \ldots, n_k \to \infty$ on a

$$P\{\Delta < z\} \sim \mathrm{k}_k(z) = \frac{4}{\Gamma\left(\frac{k-1}{2}\right)(2z^2)^{\frac{k-1}{2}}} \times \\ \times \sum_{s=1}^{\infty} \frac{\mu_s^{k-3}}{[J'_{\frac{k-3}{2}}(\mu_s)]^2} \exp\left[-\frac{\mu_s^2}{2z^2}\right], \quad z > 0, \tag{5}$$

où μ_s est le s-ième zéro positif de la fonction de Bessel $J_{\frac{k-3}{2}}(z)$. Vérifier que pour $k = 2$ on a

$$\mathrm{k}_2(z) = \mathrm{k}(z) = \frac{\sqrt{2\pi}}{z} \sum_{s=1}^{\infty} \exp\left[-\frac{\pi^2(2s-1)^2}{8z^2}\right], \quad z > 0. \tag{6}$$

2.4. Soit m le nombre de réalisations d'un événement lors de n essais indépendants, pour chacun desquels la probabilité de cet événement est égale à p. En utilisant l'inégalité de Tchébychev montrer que pour un $\varepsilon > 0$ quelconque on a

$$\lim_{n\to\infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1. \tag{7}$$

2.5. Supposons qu'il y a lieu d'estimer une fonction linéaire de paramètres inconnus a_k, $k = 1, \ldots, m$, de la forme

$$a = \sum_{k=1}^{m} g_k a_k \tag{8}$$

(g_k sont des grandeurs connues) d'après les échantillons de N variables aléatoires $x_1, \ldots, x_N$ dont les moyennes sont

$$m_1\{x_i\} = \sum_{k=1}^{m} f_{ik} a_k, \quad i = 1, \ldots, N, \tag{9}$$

et les variances sont limitées (elles sont connues à une constante près, c'est-à-dire $M_2\{x_i\} = \frac{\sigma^2}{\lambda_i}$ où σ^2 est inconnu, et λ_i connus) (cf. [7]).

Montrer qu'une estimation de la forme

$$\hat{a} = \sum_{k=1}^{m} g_k \hat{a}_k, \tag{10}$$

où $\hat{a}_k$ est l'estimation du paramètre a_k obtenue par la méthode des moindres carrés (voir § 2.5.4), dans la classe des estimations linéaires non biaisées, a une variance minimale (même si les variables aléatoires $x_1, \ldots, x_N$ ne sont pas réparties suivant une loi normale).

2.6. Connaissant la variance σ^2 d'une variable aléatoire normale dont la moyenne a est répartie suivant la loi exponentielle

$$w_1(a) = \frac{1}{a_0} e^{-\frac{a}{a_0}}, \quad a > 0, \quad a_0 > 0, \tag{11}$$

montrer que l'estimation de Bayes du paramètre a, pour une fonction de pertes simple, est égale à

$$\hat{a}_s = \frac{1}{n} \sum_{k=1}^{n} x_k - \frac{\sigma^2}{n a_0}, \tag{12}$$

et pour une fonction de pertes quadratique, à

$$\hat{a}_q = \hat{a}_s + \frac{\sigma}{\sqrt{2\pi}} \left[F\left(\frac{\hat{a}_s}{\sigma}\right) \right]^{-1} e^{-\frac{\hat{a}_s^2}{2\sigma^2}}, \tag{13}$$

où $F(x)$ est la fonction de Laplace. Vérifier que l'estimation (12) pour une fonction de pertes simple correspond à l'estimation du maximum de densité de probabilité a posteriori du paramètre a, et pour $n \to \infty$ tend asymptotiquement vers l'estimation du maximum de vraisemblance, tandis que l'estimation (13) pour une fonction de pertes quadratique ne présente pas cette particularité.

Chapitre 3

STATISTIQUE DES PROCESSUS ALÉATOIRES

3.1. DEUX MÉTHODES DE REPRÉSENTATION D'UN PROCESSUS ALÉATOIRE

Un processus aléatoire peut être défini de deux manières différentes (cf. § 4.1.2, tome I). Si pour caractériser les propriétés probabilistes du processus aléatoire $\xi(t)$ on se limite à une fonction de répartition $F_N(x_1, \ldots, x_N, t_1, \ldots, t_N)$ de dimension finie, le processus aléatoire se trouve identifié à un ensemble de N *variables aléatoires* $\xi_1 = \xi(t_1), \ldots, \xi_N = \xi(t_N)$ qui dans le cas général sont un ensemble de variables aléatoires *dépendantes*. Le choix des instants $t_1, \ldots, t_N$ est aléatoire. Si, par exemple, on se donne l'intervalle d'observation $(-T, T)$, on peut poser $t_k = \frac{2k-N}{N} T$, $k = 1, \ldots$ $\ldots, N$. Le résultat de l'observation du processus aléatoire est dans le cas envisagé un échantillon prélevé sur la répartition multidimensionnelle F_N (cf. § 2.7.1).

Une autre méthode concervant toute l'information probabiliste sur le processus aléatoire consiste à considérer l'ensemble de ses *réalisations* comme fonctions du temps t. La mesure probabiliste déterminée sur l'ensemble des réalisations est la fonctionnelle caractéristique du processus aléatoire (cf. (4.9), tome I)

$$\Theta[v(t)] = m_1 \left\{ \exp\left[i \int_{-\infty}^{\infty} v(t)\, \xi(t)\, dt \right] \right\} ,$$

ici $v(t)$ est une fonction continue et l'intégrale converge en moyenne quadratique. Dans cette méthode, le résultat de l'observation du processus aléatoire est un nombre fini de ses réalisations tronquées

$$\xi_T^{(k)}(t) = \begin{cases} \xi^{(k)}(t), & |t| \leqslant T, \\ 0, & |t| > T, \end{cases} \qquad k = 1, 2, \ldots, n.$$

Dans de nombreux cas, pour obtenir des conclusions statistiques sur les caractéristiques d'un processus aléatoire, il suffit (quand c'est possible) d'écrire le processus aléatoire comme la somme de processus aléatoires *quasi déterministes*. Les paramètres aléatoires (ne dépendant pas du temps) des composantes sont appelés parfois *coordonnées* du processus aléatoire. Le choix de ces coordonnées n'est pas univoque et dépend des conditions du problème.

Deux types de coordonnées souvent utilisées dans les applications techniques des processus aléatoires peuvent être envisagés, à savoir les valeurs échantillonnées des processus à spectre énergétique limité et les coordonnées non corrélées.

3.2. ÉCHANTILLONNAGE À DES INSTANTS DISCRETS

3.2.1. Théorème de Kotelnikov (direct). Soit une fonction déterministe $f(t)$

$$f(t)=\frac{1}{2\pi}\int_{-\Delta}^{\Delta} Z(\omega)\,e^{i\omega t}\,d\omega \qquad (3.1)$$

dont le spectre (transformée de Fourier) $Z(\omega)$ est continu et limité à la bande $(-\Delta, \Delta)$, c'est-à-dire est une fonction de fréquence, continue, à support borné, égale à zéro pour $|\omega| \geqslant \Delta$.

La fonction à support borné $Z(\omega)$ peut être écrite sous la forme d'une série de Fourier sur l'intervalle $(-\Omega, \Omega)$, à condition que $\Omega \geqslant \Delta$, soit:

$$Z(\omega)=\sum_{n=-\infty}^{\infty} c_n e^{-\frac{2\pi i n}{2\Omega}\omega}, \qquad (3.2)$$

où

$$c_n=\frac{1}{2\Omega}\int_{-\Omega}^{\Omega} Z(\omega)\,e^{\frac{2\pi i n}{2\Omega}\omega}\,d\omega=\frac{1}{2\Omega}\int_{-\Delta}^{\Delta} Z(\omega)\,e^{\frac{2\pi i n}{2\Omega}\omega}\,d\omega=\frac{2\pi}{2\Omega} f\left(\frac{\pi n}{\Omega}\right). \qquad (3.2')$$

La série se trouvant dans le second membre de (3.2) est une fonction de la fréquence de période 2Ω ne coïncidant avec $Z(\omega)$ que sur l'intervalle principal de périodicité $(-\Omega, \Omega)$ et différant sur les autres intervalles $(\pm k\Omega, \pm(k+1)\Omega)$, $k \geqslant 1$. En portant (3.2) dans (3.1) et compte tenu de (3.2') on obtient

$$f(t)=\sum_{n=-\infty}^{\infty} f\left(\frac{\pi n}{\Omega}\right)\frac{\sin\left[\Omega\left(t-\frac{\pi n}{\Omega}\right)\right]}{\Omega\left(t-\frac{\pi n}{\Omega}\right)}, \quad \Omega \geqslant \Delta. \qquad (3.3)$$

L'expression (3.3) est une formule d'*interpolation* permettant de rétablir *exactement* toutes les valeurs de $f(t)$ sur l'axe des temps de $-\infty$ à $+\infty$ d'après les échantillons de la fonction $f(t)$ faits à des instants discrets séparés de $2\Omega \geqslant 2\Delta$.

Pour la fréquence d'échantillonnage minimale possible $2\Omega = 2\Delta$ on a

$$f(t)=\sum_{n=-\infty}^{\infty} f\left(\frac{\pi n}{\Delta}\right)\frac{\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi n}{\Delta}\right)}. \qquad (3.4)$$

La formule (3.4) est l'expression du théorème connu de Kotelnikov en vertu duquel une fonction déterministe $f(t)$ de spectre limité est entièrement donnée par ses valeurs discrètes en des points disposés les uns des autres à une distance égale à $\frac{2\pi}{2\Delta}$ *) où Δ est la pulsation maximale dans le spectre de la fonction $f(t)$.

En vertu de (3.4), on peut obtenir $f(t)$ par superposition d'une suite de fonctions de la forme $\frac{\sin \Delta t}{\Delta t}$, obtenues par déplacement sur l'axe des temps par rapport à la précédente d'une grandeur égale à l'intervalle d'échantillonnage et multiplication par la valeur de lecture (d'échantillonnage) f. Il est facile de voir que chacune des fonctions $\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]/\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]$ a un maximum égal à l'unité au point $t=\frac{\pi n}{\Delta}$, et en tous les autres points $t=\frac{\pi k}{\Delta}$, $k\neq n$ (k étant un nombre entier quelconque, y compris zéro) elle est nulle.

Notons que, le spectre de la fonction $f(t)$ étant limité, celle-ci ne peut être à support borné, c'est-à-dire que l'on ne peut donner un intervalle de valeurs (même pour des t très grands) où la fonction soit identiquement nulle.

Si la fonction $f(t)$ n'est donnée que sur un intervalle de temps fini $(0, T)$ où l'on a N intervalles d'échantillonnage, c'est-à-dire si $\frac{N\pi}{\Delta}\leqslant T<\frac{(N+1)\pi}{\Delta}$, en définissant cette fonction à l'extérieur de l'intervalle $(0, T)$ de telle sorte que pour $n<0$ et $n>N$ on ait $f\left(\frac{\pi n}{\Delta}\right)=0$, on obtient dans ce cas à partir de (3.4)

$$f(t)=\sum_{n=1}^{N} f\left(\frac{\pi n}{\Delta}\right)\frac{\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi n}{\Delta}\right)},\qquad N=\left[\frac{2\Delta T}{2\pi}\right], \tag{3.5}$$

où [ ] donne la plus grande partie entière du nombre se trouvant entre crochets. Pour $N\gg 1$ on peut sans erreur importante poser $N=\frac{2\Delta T}{2\pi}$.

Les échantillons de $f(t)$ permettent également de définir la valeur de l'intégrale du carré de cette fonction

$$E=\int_{-\infty}^{\infty} f^2(t)\,dt \tag{3.6}$$

*) Notons qu'à $\frac{\pi n}{\Delta}$ dans (3.4) on peut substituer $t_0+\frac{\pi n}{\Delta}$, t_0 étant un nombre donné quelconque (cf. le problème 3.4).

en supposant que E soit borné. Portant (3.4) dans (3.6) et compte tenu de ce que le système de fonctions

$$\frac{\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi n}{\Delta}\right)}, \quad n=0, \pm 1, \ldots,$$

est *orthogonal* sur l'intervalle $(-\infty, \infty)$ pour le poids $\varphi(t) \equiv 1$ et que la norme des fonctions est $\frac{\pi}{\Delta}$ (cf. annexe IV), on obtient ainsi

$$E=\int_{-\infty}^{\infty} f^2(t)\,dt=\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} f\left(\frac{\pi n}{\Delta}\right) f\left(\frac{\pi k}{\Delta}\right)\times$$

$$\times\int_{-\infty}^{\infty}\frac{\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi n}{\Delta}\right)}\frac{\sin\left[\Delta\left(t-\frac{\pi k}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi k}{\Delta}\right)}\,dt=\frac{\pi}{\Delta}\sum_{n=-\infty}^{\infty} f^2\left(\frac{\pi n}{\Delta}\right). \tag{3.7}$$

Si les valeurs échantillonnées de $f\left(\frac{\pi n}{\Delta}\right)$ sont égales à zéro pour $n<0$ et $n>N$ pour $N \gg 1$ on a

$$E=\frac{\pi}{\Delta}\sum_{n=0}^{\frac{\Delta T}{\pi}} f^2\left(\frac{\pi n}{\Delta}\right). \tag{3.7'}$$

Notons également que les formules (3.5) et (3.7′) peuvent servir d'approximation pour le cas où la grandeur

$$1-\frac{\pi}{E\Delta}\sum_{n=0}^{\frac{\Delta T}{\pi}} f^2\left(\frac{\pi n}{\Delta}\right)$$

est négligeable.

Notons enfin que, avec des restrictions supplémentaires, la formule (3.4) reste vraie également dans le cas où le spectre de la fonction $f(t)$ se compose d'une partie continue occupant une certaine bande de fréquence et d'une partie discrète formée par une somme de fonctions delta (cf. § 4.2.5 et 4.2.7, tome I)

$$Z(\omega)=Z_c(\omega)+a_0\delta(\omega)+\sum_k a_k\left[\delta(\omega+\omega_k)+\delta(\omega-\omega_k)\right]. \tag{3.8}$$

En vertu de (3.8), la fonction $f(t)$ contient une composante constante a_0 différente de zéro et des composantes harmoniques de périodes $\frac{2\pi}{\omega_k}$. C'est pourquoi la formule de Kotelnikov (3.4), pour des

fonctions de spectre (3.8), s'écrit comme suit:

$$f(t)-a_0=\sum_{n=-\infty}^{\infty}\left[f\left(\frac{\pi n}{\Delta}\right)-a_0\right]\frac{\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi n}{\Delta}\right)}, \quad (3.9)$$

de plus $\frac{n}{2\Delta}\neq\frac{1}{\omega_k}$ pour tout k, car sans cela la série du second membre de (3.9) serait divergente pour $t=\frac{\pi n}{\Delta}$.

A titre d'exemple d'application de la formule de Kotelnikov mentionnons la formule d'interpolation pour la *fonction de corrélation* $B(\tau)$ d'un processus aléatoire stationnaire au sens général, dont le spectre énergétique $F(\omega)$ est continu et identiquement nul pour $|\omega|>\Delta$:

$$B(\tau)=\sum_{n=-\infty}^{\infty}B\left(\frac{\pi n}{\Delta}\right)\frac{\sin\left[\Delta\left(\tau-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(\tau-\frac{\pi n}{\Delta}\right)}. \quad (3.10)$$

En vertu de (3.3), dans le second membre de (3.10) on peut substituer à Δ un $\Omega\geqslant\Delta$ quelconque. En utilisant le théorème de Wiener-Khintchine (cf. § 4.2.3, tome I) et compte tenu de la relation

$$\int_{-\infty}^{\infty}\frac{\sin\left[\Delta\left(\tau-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(\tau-\frac{\pi n}{\Delta}\right)}e^{-i\omega\tau}\,d\tau=\begin{cases}\frac{\pi}{\Delta}e^{-\frac{i\pi n\omega}{\Delta}}, & |\omega|\leqslant\Delta,\\ 0, & |\omega|>\Delta,\end{cases} \quad (3.11)$$

on déduit immédiatement de (3.10) la relation entre le spectre énergétique limité et les échantillons de la fonction de corrélation, soit:

$$F(\omega)=\frac{2\pi}{\Delta}\sum_{n=-\infty}^{\infty}B\left(\frac{\pi n}{\Delta}\right)e^{-\frac{i\pi n\omega}{\Delta}}=$$

$$=\frac{2\pi}{\Delta}\left[B(0)+2\sum_{n=1}^{\infty}B\left(\frac{\pi n}{\Delta}\right)\cos\frac{\pi n\omega}{\Delta}\right],\quad |\omega|<\Delta. \quad (3.12)$$

Pour un bruit blanc ayant traversé un filtre parfait passe-bas (cf. (5.31), tome I) la fonction de corrélation est nulle aux points $\tau=\frac{\pi n}{\Delta}$ pour $n>0$. Dans les formules (3.10) et (3.12) seuls les premiers termes restent et celles-ci deviennent les formules (5.44) et (5.45) du premier tome (pour $\omega_0=0$). On voit également qu'en vertu de (3.10) et (3.12) un processus stationnaire au sens général, dont la fonction de corrélation s'annule pour $\tau=\frac{\pi n}{\Delta}$, est équivalent par ses caractéristiques énergétiques à un bruit blanc après filtrage parfait.

3.2.2. Théorème de Kotelnikov (inverse). Soit maintenant $f(t)$ une fonction continue, identiquement nulle pour $|t| \geqslant \frac{T}{2}$.

Le spectre de cette fonction (sa transformée de Fourier) est égal à

$$Z(\omega) = \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\, e^{-i\omega t}\, dt. \tag{3.13}$$

Cependant, le spectre continu $Z(\omega)$ de la fonction $f(t)$ limitée dans le temps par l'intervalle $|t| \leqslant \frac{T}{2}$ est défini d'une manière univoque par les densités spectrales aux pulsations $\omega_k = \frac{2\pi k}{T}$, où k est un nombre entier quelconque (y compris zéro). En effet, en développant en série de Fourier sur l'intervalle $|t| \leqslant \frac{T}{2}$ la fonction $f(t)$ à support borné, on obtient *)

$$f(t) = \sum_{n=-\infty}^{\infty} a_n e^{\frac{2\pi i n}{T} t}, \tag{3.14}$$

où

$$a_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\, e^{-\frac{2\pi i n}{T} t}\, dt = \frac{1}{T} Z\left(\frac{2\pi i n}{T}\right). \tag{3.14'}$$

En portant (3.14) dans (3.13) et en utilisant (3.14′), on obtient après quelques réductions

$$Z(\omega) = \sum_{n=-\infty}^{\infty} Z\left(\frac{2\pi i n}{T}\right) \frac{\sin\left(\frac{\omega T}{2} - \pi n\right)}{\frac{\omega T}{2} - \pi n}. \tag{3.15}$$

La propriété exprimée par la formule (3.15) du spectre d'une fonction bornée dans le temps est analogue à la propriété d'une fonction à spectre limité, donnée par le théorème de Kotelnikov [cf. (3.4)].

Notons que le développement d'une fonction $f(t)$ à support borné sur l'intervalle $\left(-\frac{T}{2}, \frac{T}{2}\right)$ en série de Fourier (3.14) définit $f(t)$ en fonction des valeurs de son spectre $Z\left(\frac{2\pi i n}{T}\right)$ sur des pulsations distantes les unes des autres de $\frac{2\pi}{T}$.

---

*) Comme au § 3.2.1, on peut obtenir une relation plus générale donnant un nombre superflu d'échantillons de $Z(\omega)$ en développant $f(t)$ en série de Fourier sur l'intervalle $|t| < \frac{T_1}{2}$ pour $T_1 \geqslant T$.

Il est également facile de trouver l'équivalent de la formule (3.7)

$$E=\int_{-\frac{T}{2}}^{\frac{T}{2}} f^2(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}|Z(\omega)|^2\,d\omega=\frac{1}{T}\sum_{n=-\infty}^{\infty}\left|Z\left(\frac{2\pi in}{T}\right)\right|^2 \qquad (3.16)$$

ou, compte tenu de (3.14'),

$$\frac{E}{T}=\sum_{n=-\infty}^{\infty} a_n^2, \qquad (3.16')$$

comme on pouvait s'y attendre en conformité du théorème de Parseval (cf., par exemple, [8]).

Pour illustrer le théorème inverse de Kotelnikov considérons la formule d'interpolation pour un spectre énergétique continu $F(\omega)$ d'un processus aléatoire stationnaire au sens général dont la fonction de corrélation $B(\tau)$ est identiquement nulle pour $|\tau|\geqslant\tau_0$:

$$F(\omega)=\sum_{n=-\infty}^{\infty} F\left(\frac{\pi n}{\tau_0}\right)\frac{\sin(\omega\tau_0-\pi n)}{\omega\tau_0-\pi n}. \qquad (3.17)$$

En utilisant le théorème de Wiener-Khintchine (cf. § 4.2.3, tome I) et compte tenu de (3.11) on obtient à partir de (3.17) la relation entre la fonction de corrélation à support borné et les valeurs échantillonnées du spectre énergétique:

$$B(\tau)=\frac{1}{4\tau_0}\sum_{n=-\infty}^{\infty} F\left(\frac{\pi n}{\tau_0}\right)e^{\frac{i\pi n\tau}{\tau_0}}=$$

$$=\frac{1}{4\tau_0}\left[F(0)+2\sum_{n=1}^{\infty}F\left(\frac{\pi n}{\tau_0}\right)\cos\frac{\pi n\tau}{\tau_0}\right],\quad |\tau|<\tau_0. \qquad (3.18)$$

3.2.3. Extension du théorème de Kotelnikov aux processus aléatoires. Soit $\xi(t)$ un processus aléatoire continu en moyenne quadratique et stationnaire au sens général, dont le spectre énergétique $F_\xi(\omega)$ est continu et identiquement nul à l'extérieur de la bande de fréquence $|\omega|<\Delta$. Montrons que pour ce processus l'égalité suivante est vérifiée (en moyenne quadratique, cf. § 3.5, tome I):

$$\xi(t)=\sum_{n=-\infty}^{\infty}\xi\left(\frac{\pi n}{\Delta}\right)\frac{\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi n}{\Delta}\right)} \qquad (3.19)$$

L'égalité (3.19) qui généralise le théorème de Kotelnikov aux processus aléatoires exprime le fait qu'un processus continu en moyenne quadratique, à spectre énergétique limité, se trouve être entière-

ment défini par un ensemble dénombrable de variables aléatoires (coordonnées du processus aléatoire)

$$\xi_n = \xi\left(\frac{\pi n}{\Delta}\right), \quad n = 0, \ \pm 1, \ \ldots \qquad (3.19')$$

Avec les restrictions mentionnées, le processus aléatoire est une somme de processus quasi déterministes du type

$$\xi_n \frac{\sin\left[\Delta\left(t - \frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t - \frac{\pi n}{\Delta}\right)}.$$

Pour démontrer que l'égalité (3.19) est vraie en moyenne quadratique, il y a lieu d'établir l'égalité des fonctions de corrélation des processus figurant dans les deux membres de l'égalité. On a

$$m_1\left\{\sum_{r=-\infty}^{\infty} \xi\left(\frac{\pi r}{\Delta}\right) \frac{\sin\left[\Delta\left(t - \frac{\pi r}{\Delta}\right)\right]}{\Delta\left(t - \frac{\pi r}{\Delta}\right)} \times \right.$$

$$\left. \times \sum_{k=-\infty}^{\infty} \xi\left(\frac{\pi k}{\Delta}\right) \frac{\sin\left[\Delta\left(t + \tau - \frac{\pi k}{\Delta}\right)\right]}{\Delta\left(t + \tau - \frac{\pi k}{\Delta}\right)}\right\} =$$

$$= \sum_{r=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} m_1\left\{\xi\left(\frac{\pi r}{\Delta}\right) \xi\left(\frac{\pi k}{\Delta}\right)\right\} \frac{\sin\left[\Delta\left(t - \frac{\pi r}{\Delta}\right)\right]}{\Delta\left(t - \frac{\pi r}{\Delta}\right)} \times$$

$$\times \frac{\sin\left[\Delta\left(t + \tau - \frac{\pi k}{\Delta}\right)\right]}{\Delta\left(t + \tau - \frac{\pi k}{\Delta}\right)}.$$

Comme $\xi(t)$ est stationnaire au sens général, on a

$$m_1\left\{\xi\left(\frac{\pi r}{\Delta}\right) \xi\left(\frac{\pi k}{\Delta}\right)\right\} = B_\xi\left[\frac{\pi}{\Delta}(r - k)\right].$$

De plus, pour des t et τ arbitraires, la somme double devient une somme simple *), d'où

$$\sum_{r=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} B_\xi\left[\frac{\pi}{\Delta}(r - k)\right] \frac{\sin\left[\Delta\left(t - \frac{\pi r}{\Delta}\right)\right]}{\Delta\left(t - \frac{\pi r}{\Delta}\right)} \times$$

$$\times \frac{\sin\left[\Delta\left(t + \tau - \frac{\pi k}{\Delta}\right)\right]}{\Delta\left(t + \tau - \frac{\pi k}{\Delta}\right)} = \sum_{n=-\infty}^{\infty} B_\xi\left(\frac{\pi n}{\Delta}\right) \frac{\sin\left[\Delta\left(\tau - \frac{\pi n}{\Delta}\right)\right]}{\Delta\left(\tau - \frac{\pi n}{\Delta}\right)}. \qquad (3.20)$$

*) Ceci découle immédiatement de la formule (6) du problème 3.3.

Comme le spectre énergétique $F_\xi(\omega)$ n'est défini que dans la bande de fréquence $|\omega| \leqslant \Delta$, en vertu de (3.10) la somme du second membre de (3.20) coïncide avec la fonction de corrélation $B_\xi(\tau)$.

En utilisant les remarques faites au § 3.2.1, on peut étendre quelque peu les conditions d'applicabilité de la formule (3.19). Notons avant tout que cette formule est vraie lorsque l'on remplace Δ par un $\Omega \geqslant \Delta$ quelconque [cf. (3.3)], de même si l'on remplace dans le second membre $\frac{\pi n}{\Delta}$ par $t_0 + \frac{\pi n}{\Delta}$ (problème 3.4). De plus, on peut lever la restriction concernant la continuité du spectre énergétique du processus $\xi(t)$ et admettre que le spectre contienne une partie discrète (somme de fonctions delta) aux pulsations ω_k à condition que $\Delta \neq n\omega_k$. Enfin, on peut généraliser (3.19) aux processus aléatoires à bande étroite (cf. problème 3.1).

Les coordonnées du processus aléatoire $\xi(t)$ choisies conformément au théorème de Kotelnikov sont en général des variables aléatoires *corrélées*. D'autre part, l'utilisation de ces coordonnées est limitée par les conditions exigeant que le processus soit stationnaire, au moins au sens général, et que son spectre énergétique soit borné dans le temps. Nous allons maintenant passer à l'étude des coordonnées d'un processus aléatoire non astreint à de telles restrictions.

3.3. DÉVELOPPEMENT ORTHOGONAL D'UN PROCESSUS ALÉATOIRE

3.3.1. Développement orthogonal d'une fonction de corrélation. L'étude des propriétés générales des fonctions de corrélation des processus aléatoires montre que toute fonction de corrélation $B(t, \tau)$ est un noyau symétrique d'une forme quadratique semi-définie positive (cf. § 4.2, tome I), c'est-à-dire que pour tous $n, u_1, \ldots, u_n$ on a

$$\sum_{i=1}^{n} \sum_{j=1}^{n} B(t_i, t_j)\, u_i u_j \geqslant 0 \tag{3.21}$$

ou pour une fonction $f(t)$ quelconque si toutefois $\int_{-T}^{T} f^2(t)\, dt < \infty$, on a

$$\int_{-T}^{T} \int_{-T}^{T} B(t, y) f(t) f(y)\, dt\, dy \geqslant 0. \tag{3.21'}$$

En utilisant (3.21) pour (3.21'), il est possible de représenter la fonction de corrélation d'un processus aléatoire sous la forme de la série suivante:

$$B(t, y) = \sum_{k=1}^{\infty} \frac{\varphi_k(t)\, \varphi_k(y)}{\lambda_k}, \tag{3.22}$$

où $\varphi_k(t)$ et λ_k sont les fonctions propres (solutions) et les nombres caractéristiques de l'équation intégrale linéaire homogène

$$\varphi(t) = \lambda \int_{-T}^{T} B(t, y) \varphi(y) dy, \qquad |t| \leqslant T. \tag{3.23}$$

Les fonctions propres $\varphi_k(t)$ forment un système de fonctions orthogonales normées, c'est-à-dire

$$\int_{-T}^{T} \varphi_k(t) \varphi_j(t) dt = \begin{cases} 0, & k \neq j, \\ 1, & k = j. \end{cases} \tag{3.23'}$$

En multipliant les deux membres de (3.23) par $\varphi(t)$, en intégrant ensuite sur t de $-T$ à T et compte tenu de (3.21) et (3.23'), on peut voir que les nombres caractéristiques λ_k sont positifs. De plus, si la fonction de corrélation $B(t, y)$ est définie positive [c'est-à-dire que dans (3.21) et (3.21') l'égalité à zéro se trouve exclue] on peut montrer que l'ensemble des fonctions propres est achevé. Ceci signifie que sur l'intervalle $|t| \leqslant T$ il n'existe pas de fonction $\psi(t)$ qui soit orthogonale à tous les $\varphi_k(t)$.

Pour la fonction de corrélation $B(\tau)$ d'un processus aléatoire stationnaire au sens général, le développement (3.22) peut s'écrire comme suit :

$$B(t-y) = \sum_{k=1}^{\infty} \frac{\varphi_k(t) \varphi_k(y)}{\lambda_k}, \tag{3.24}$$

pour $t = y$ on a

$$B(0) = \sum_{k=1}^{\infty} \frac{\varphi_k^2(t)}{\lambda_k}. \tag{3.24'}$$

En intégrant les deux membres de (3.24') par rapport à t, on obtient compte tenu de (3.23')

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_k} = 2TB(0),$$

c'est-à-dire que la somme de grandeurs inversement proportionnelles aux nombres caractéristiques est égale à l'énergie moyenne du processus sur l'intervalle $(-T, T)$.

Comme on peut le voir facilement, dans le cas général la somme des puissances inverses des nombres caractéristiques peut s'exprimer comme suit [cf. (3.32), tome I]

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_k^n} = \int_{-T}^{T} B^{(n)}(u, v) du, \tag{3.24''}$$

où $B^{(n)}(u, v)$ est l'itération d'ordre n de la fonction de corrélation $B(t, y)$

$$B^{(n)}(u, v)=\int_{-T}^{T}\ldots\int_{-T}^{T} B(u, t_1) B(t_1, t_2)\ldots B(t_{n-1}, v)\, dt_1\ldots dt_{n-1},\ n\geqslant 2,$$
$$B^{(1)}(u, v)\equiv B(u, v).$$

Notons que pour la fonction de corrélation d'un bruit blanc d'intensité unitaire on a $B(t, y) = \delta(t - y)$ et compte tenu de (3.23) tous les nombres caractéristiques sont égaux entre eux et égaux à l'unité : $\lambda = 1$. On a alors en vertu de (3.22)

$$\delta(t-y)=\sum_{k=1}^{\infty}\varphi_k(t)\,\varphi_k(y), \qquad (3.24''')$$

où $\{\varphi_k(t)\}$ est un système quelconque de fonctions orthonormées.

3.3.2. Coordonnées non corrélées d'un processus aléatoire. Soit un processus aléatoire $\xi(t)$ continu en moyenne quadratique, de moyenne nulle et dont la fonction de corrélation est $B(t, y)$. Nous allons prendre pour *coordonnées* du processus aléatoire $\xi(t)$ les variables aléatoires (l'intégrale en moyenne quadratique)

$$\xi_k=\sqrt{\lambda_k}\int_{-T}^{T}\xi(t)\,\varphi_k(t)\,dt, \qquad (3.25)$$

où $\varphi_k(t)$ et λ_k sont les fonctions propres et les nombres caractéristiques de l'équation intégrale (3.23). Les valeurs moyennes de ces variables aléatoires sont, de toute évidence, nulles. De plus elles ne sont pas corrélées deux à deux et ont toutes des variances égales à l'unité, car en vertu de (3.23) et (3.23′), on a

$$m_1\{\xi_k\xi_m\}=\sqrt{\lambda_k\lambda_m}\int_{-T}^{T}\int_{-T}^{T}B(t, y)\,\varphi_k(t)\,\varphi_k(y)\,dt\,dy=$$
$$=\sqrt{\frac{\lambda_m}{\lambda_k}}\int_{-T}^{T}\varphi_k(y)\,\varphi_m(y)\,dy=\begin{cases}0, & k\neq m,\\ 1, & k=m.\end{cases} \qquad (3.26)$$

Montrons maintenant que pour un t quelconque appartenant à l'intervalle $(-T, T)$ on a en moyenne quadratique

$$\xi(t)=\sum_{k=1}^{\infty}\xi_k\frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \qquad (3.27)$$

c'est-à-dire que

$$\xi_N(t)=\sum_{k=1}^{N}\xi_k\frac{\varphi_k(t)}{\sqrt{\lambda_k}} \qquad (3.27')$$

converge en moyenne quadratique vers $\xi(t)$ pour $N \to \infty$.

Compte tenu de (3.25), (3.26) et (3.27') on trouve

$$m_1\{\xi(t)\,\xi_N(t)\} = m_1\left\{\xi(t) \sum_{k=1}^{N} \varphi_k(t) \int_{-T}^{T} \xi(u)\,\varphi_k(u)\,du\right\} =$$

$$= \sum_{k=1}^{N} \int_{-T}^{T} B(t, u)\,\varphi_k(u)\,\varphi_k(t)\,du = \sum_{k=1}^{N} \frac{\varphi_k^2(t)}{\lambda_k},$$

$$m_1\{\xi_N^2(t)\} = \sum_{k=1}^{N} \sum_{n=1}^{N} \frac{\varphi_k(t)\,\varphi_n(t)}{\sqrt{\lambda_k \lambda_n}}\, m_1\{\xi_k \xi_n\} = \sum_{k=1}^{N} \frac{\varphi_k^2(t)}{\lambda_k},$$

d'où

$$m_1\{[\xi(t) - \xi_N(t)]^2\} = B(t, t) - \sum_{k=1}^{N} \frac{\varphi_k^2(t)}{\lambda_k}. \tag{3.28}$$

A partir de (3.22) pour $t = y$ on obtient

$$B(t, t) = \sum_{k=1}^{\infty} \frac{\varphi_k^2(t)}{\lambda_k}, \tag{3.29}$$

et, par conséquent, pour $N \to \infty$ le second membre de (3.28) tend vers zéro, c'est-à-dire que $\xi_N(t)$ converge en moyenne quadratique vers $\xi(t)$.

Ainsi se trouve démontrée la possibilité de représenter un processus aléatoire par une somme de processus aléatoires quasi déterministes du type $\xi_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}$, où $\varphi_k(t)$ et λ_k sont déterminés par la fonction de corrélation du processus, et ξ_k sont des variables aléatoires obtenues par intégration « pondérée » du processus aléatoire [cf. (3.25)].

Quand la moyenne de $\xi(t)$ n'est pas nulle mais égale à $a(t)$, il y a lieu d'utiliser le développement (3.27) pour trouver l'écart d'un processus aléatoire de sa valeur moyenne, on a alors

$$\xi(t) = a(t) + \sum_{k=1}^{\infty} \xi_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \quad |t| \leqslant T, \tag{3.30}$$

ici le noyau de l'équation intégrale (3.23) n'est plus $B(t, y)$ mais $B(t, y) - a(t)\,a(y)$.

Notons que l'on peut représenter la fonction déterministe $a(t)$ sur l'intervalle $|t| \leqslant T$ sous la forme d'un développement orthogonal, soit:

$$a(t) = \sum_{k=1}^{\infty} a_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \quad |t| \leqslant T, \tag{3.31}$$

où

$$a_k = \sqrt{\lambda_k} \int_{-T}^{T} a(u)\, \varphi_k(u)\, du, \tag{3.32}$$

$\varphi_k(t)$ et λ_k étant un ensemble *quelconque* de fonctions propres et de nombres caractéristiques *). On peut vérifier (3.31) en y portant (3.32) et en utilisant (3.24").

Réunissant (3.30) et (3.31) on peut écrire le développement du processus aléatoire $\xi(t)$ de moyenne non nulle sous la forme suivante:

$$\xi(t) = \sum_{k=1}^{\infty} (\xi_k + a_k) \frac{\varphi_k(t)}{\sqrt{\lambda_k}}. \tag{3.33}$$

Notons qu'il est possible d'obtenir le développement d'un processus aléatoire suivant un ensemble quelconque de fonctions déterministes orthogonales **), mais dans ce cas général les coordonnées du processus seront corrélées. Ce n'est que pour un choix judicieux de ces fonctions, compte tenu des propriétés corrélationnelles du processus [cf. (3.23)], que les coordonnées deviennent non corrélées. Si, de plus, le processus $\xi(t)$ est normal, ces coordonnées non corrélées deviennent des variables aléatoires *indépendantes*.

La difficulté majeure que l'on rencontre lors de l'utilisation pratique du développement orthogonal du processus aléatoire (3.27) ou (3.30), c'est-à-dire lorsqu'on le représente sous la forme d'un ensemble dénombrable de variables aléatoires, est de trouver les solutions exactes de l'équation intégrale (3.23). Ce n'est que pour des processus stationnaires dont le spectre est une fraction rationnelle qu'il est possible de trouver ces solutions par intégration d'une équation différentielle linéaire (cf., par exemple, annexe 2 dans [3]).

3.3.3. Développement orthogonal d'un processus aléatoire complexe. Dans certains problèmes on a besoin de généraliser le développement orthogonal mentionné ci-dessus au cas d'un processus aléatoire complexe (cf. § 4.7.1, tome I)

$$\zeta(t) = \xi(t) + i\eta(t).$$

---

*) On suppose également que $\int_{-T}^{T} a^2(t)\, dt = \sum_{k=1}^{\infty} \frac{a_k^2}{\lambda_k} < \infty$.

**) Par exemple, les fonctions harmoniques $\sqrt{\frac{2}{T}} \cos k\omega_0 t$, $\sqrt{\frac{2}{T}} \sin k\omega_0 t$, $\omega_0 = \frac{2\pi}{T}$; dans ce cas le développement est appelé série de Fourier (en moyenne quadratique).

Tout comme dans (3.22), on peut écrire la fonction de corrélation du processus $\zeta(t)$ sous la forme suivante

$$B_\zeta(t, y) = m_1\{\zeta(t)\overline{\zeta(y)}\} = \sum_{k=1}^{\infty} \frac{\varphi_k(t)\overline{\varphi_k(y)}}{\lambda_k}, \tag{3.34}$$

où le trait indique qu'il s'agit d'une grandeur complexe conjuguée, $\varphi_k(t)$ et λ_k étant les fonctions propres et les nombres caractéristiques de l'équation intégrale (3.23) de noyau $B_\zeta(t, y)$.

Introduisant les coordonnées *complexes* non corrélées

$$\zeta_k = \sqrt{\lambda_k} \int_{-T}^{T} \zeta(t)\overline{\varphi_k(t)}\,dt, \tag{3.35}$$

$$m_1\{\zeta_k\overline{\zeta_m}\} = \begin{cases} 0, & k \neq m, \\ 1, & k = m, \end{cases} \tag{3.35'}$$

et supposant toujours que la moyenne est nulle, on obtient le développement orthogonal suivant du processus $\zeta(t)$:

$$\zeta(t) = \sum_{k=1}^{\infty} \zeta_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}. \tag{3.36}$$

Quand la moyenne $a_\zeta(t)$ du processus $\zeta(t)$ n'est pas nulle, par analogie avec (3.33) on obtient

$$\zeta(t) = \sum_{k=1}^{\infty} (\zeta_k + a_k) \frac{\varphi_k(t)}{\sqrt{\lambda_k}} \tag{3.37}$$

ou

$$a_k = \sqrt{\lambda_k} \int_{-T}^{T} a_\zeta(t)\overline{\varphi_k(t)}\,dt. \tag{3.38}$$

3.3.4. Bruit blanc à la traversée d'un filtre parfait. Pour illustrer ce qui a été dit au § 3.3.2, considérons un processus aléatoire stationnaire (au sens général) à la sortie d'un filtre passe-bas parfait, à l'entrée duquel on applique un bruit blanc de puissance N_0 par unité de bande [cf. (5.30), tome I]. Le spectre énergétique de ce processus est

$$F(\omega) = \begin{cases} 2N_0, & |\omega| \leqslant \Delta, \\ 0, & |\omega| > \Delta, \end{cases} \tag{3.39}$$

et sa fonction de corrélation

$$B(t-y) = B(0)\frac{\sin(t-y)\Delta}{(t-y)\Delta}, \quad B(0) = \frac{N_0\Delta}{2\pi}. \tag{3.39'}$$

Les fonctions propres de l'équation intégrale (3.23) de noyau (3.39′) forment un ensemble de fonctions orthonormées [cf. formule (6) du problème 3.3]

$$\varphi_k(t)=\sqrt{\frac{\Delta}{\pi}}\,\frac{\sin\Delta\left(t-\frac{\pi k}{\Delta}\right)}{\Delta\left(t-\frac{\pi k}{\Delta}\right)},\quad -\infty<t<\infty, \tag{3.40}$$

les nombres caractéristiques étant

$$\lambda_k=\frac{1}{N_0}. \tag{3.40′}$$

Les variables aléatoires

$$\xi_k=\sqrt{\frac{\Delta}{\pi N_0}}\int_{-\infty}^{\infty}\xi(t)\,\frac{\sin\left[\Delta\left(t-\frac{\pi k}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi k}{\Delta}\right)}\,dt, \tag{3.41}$$

sont en vertu de (3.25) les coordonnées non corrélées du processus tandis que son développement orthogonal

$$\xi(t)=\sqrt{\frac{\Delta}{\pi N_0}}\sum_{k=-\infty}^{\infty}\xi_k\,\frac{\sin\left[\Delta\left(t-\frac{\pi k}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi k}{\Delta}\right)}, \tag{3.41′}$$

est une série de Kotelnikov [cf. (3.19)] avec $\xi_k=\sqrt{\frac{\Delta}{\pi N_0}}\,\xi\left(\frac{\pi n}{\Delta}\right)$.

Notons que l'on peut obtenir le développement orthogonal d'un bruit blanc $\xi(t)$ d'intensité spectrale N_0 à partir d'un système quelconque de fonctions orthonormées $\{\varphi_k(t)\}$ [cf. (3.24‴)]. Les coordonnées non corrélées d'un bruit blanc sont

$$\xi_k=\frac{1}{\sqrt{N_0}}\int_{-\infty}^{\infty}\xi(t)\,\varphi_k(t)\,dt. \tag{3.42}$$

3.4. CARACTÉRISTIQUES DES COORDONNÉES OBSERVÉES D'UN PROCESSUS ALÉATOIRE

3.4.1. Fonction de vraisemblance des coordonnées observées. Supposons que le processus aléatoire $\xi(t)$ soit donné par ses coordonnées, c'est-à-dire par un ensemble fini ou dénombrable de variables aléatoires ξ_k. Si l'on se donne la durée de l'intervalle d'observation, le résultat d'une expérience peut être représenté par un élément d'un échantillon d'une distribution multidimensionnelle de l'ensemble $\xi_1, \ldots, \xi_N$. On peut représenter les résultats des observations différemment, suivant les coordonnées adoptées. Ainsi, lors de l'observation sur l'intervalle $(-T, T)$ à des intervalles de temps égaux $\frac{2T}{N}$, l'élément de l'échantillon se compose de N nombres

$x_{1k} = \xi^{(1)}(t_k)$, où $t_k = \frac{2k-N}{N}\hat{T}$, $k = 1, \ldots, N$. Si l'échantillon est prélevé dans un ensemble de variables aléatoires non corrélées (3.25), les *coordonnées observées* seront les grandeurs obtenues par intégration sur un intervalle donné de la réalisation $\xi^{(1)}(t)$ pondérée par des valeurs différentes $\sqrt{\lambda_k}\,\varphi_k(t)$, c'est-à-dire

$$x_{1k} = \sqrt{\lambda_k} \int_{-T}^{T} \xi^{(1)}(t)\,\varphi_k(t)\,dt, \quad k = 1, \ldots, N,$$

où, en vertu de ce qui a été exposé ci-dessus, $\varphi_k(t)$ et λ_k sont donnés par la fonction de corrélation du processus $\xi(t)$.

En poursuivant les observations indépendantes des coordonnées du processus aléatoire on obtient successivement les éléments de l'échantillon $x_{2k}, \ldots, x_{nk}$; $k = 1, \ldots, N$. La distribution conjointe des valeurs échantillonnées, c'est-à-dire la fonction de vraisemblance des coordonnées observées, est égale au produit des densités de probabilité multidimensionnelles des éléments de l'échantillon [cf. (2.196)]

$$L(x_{11}, \ldots, x_{nN}) = \prod_{k=1}^{n} W_N(x_{k1}, \ldots, x_{kN}). \qquad (3.43)$$

Souvent, lorsque N est suffisamment grand, on peut se limiter à un échantillon de taille $n = 1$. Notons que la limite de $W_N(x_{k1}, \ldots$ $\ldots, x_{kN})$ pour $N \to \infty$ *n'existe pas*.

3.4.2. Répartition des coordonnées observées d'un processus aléatoire normal. Nous allons maintenant obtenir les expressions de la fonction de vraisemblance des coordonnées observées d'un processus aléatoire normal, qui seront souvent utilisées dans la suite.

Pour des observations à des intervalles de temps égaux, la distribution conjointe de N coordonnées observées, donnée dans l'espace des échantillons, est une densité de probabilité normale à N dimension [cf. (4.159), tome I)

$$W_N(x_1, \ldots, x_N) = \frac{1}{(2\pi)^{N/2}\sigma_1 \ldots \sigma_N \sqrt{D}} \times$$
$$\times \exp\left\{-\frac{1}{2D}\sum_{i=1}^{N}\sum_{k=1}^{N} D_{ik}\frac{x_i - a_i}{\sigma_i}\,\frac{x_k - a_k}{\sigma_k}\right\}, \qquad (3.44)$$

où $a_k = a(t_k)$, $\sigma_k = \sigma(t_k)$ sont la moyenne et la moyenne quadratique, et D et D_{ik} le déterminant et les cofacteurs de la matrice des coefficients de corrélation.

Si l'on prend pour les coordonnées les variables aléatoires déterminées conformément à (3.25), l'expression de la fonction de vraisemblance se trouve simplifiée. En effet, si $\xi(t)$ est un processus aléatoire normal de moyenne nulle, ses coordonnées (3.25) sont des

variables aléatoires non corrélées et, par conséquent, indépendantes. Leur densité de probabilité multidimensionnelle est donc le produit de fonctions normales unidimensionnelles des variables aléatoires ξ_k de moyenne nulle et de variance unité, c'est-à-dire

$$W_N(x_1, \ldots, x_N) = \frac{1}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{N} x_k^2\right\}. \tag{3.45}$$

Pour un processus aléatoire dont la moyenne $a(t)$ n'est pas nulle, les coordonnées observées sont en vertu de (3.33) les grandeurs

$$\xi_k + a_k = \sqrt{\lambda_k} \int_{-T}^{T} [\xi(t) + a(t)]\, \varphi_k(t)\, dt, \tag{3.46}$$

on a alors

$$W_N(x_1, \ldots, x_N) = \frac{1}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{N} (x_k - a_k)^2\right\}. \tag{3.47}$$

Dans le cas où on a n observations indépendantes de la réalisation du processus sur l'intervalle $(-T, T)$, les fonctions de vraisemblance peuvent s'écrire comme suit:

$$L(x_{11}, \ldots, x_{nN}) = \frac{1}{(2\pi)^{nN/2}\sqrt{D^n}\,(\sigma_1 \ldots \sigma_N)^n} \times$$
$$\times \exp\left\{-\frac{1}{2D}\sum_{r=1}^{n}\sum_{i=1}^{N}\sum_{k=1}^{N} D_{ik}\frac{x_{ri}-a_i}{\sigma_i}\frac{x_{rk}-a_k}{\sigma_k}\right\} \tag{3.48}$$

pour un échantillonnage périodique des valeurs d'un processus aléatoire normal et

$$L(x_{11}, \ldots, x_{nN}) = \frac{1}{(2\pi)^{nN/2}} \exp\left\{-\frac{1}{2}\sum_{r=1}^{n}\sum_{k=1}^{N} (x_{rk} - a_k)^2\right\} \tag{3.49}$$

pour des coordonnées non corrélées. Pour $n = 1$ les formules (3.48) et (3.49) deviennent respectivement (3.44) et (3.47).

3.4.3. Rapport de vraisemblance et sa forme limite. Comme nous l'avons montré dans le premier chapitre, les stratégies de vérification des hypothèses statistiques concernant les paramètres de la répartition d'une variable aléatoire sont basées, pour des critères de qualité adoptés, sur le calcul du rapport de vraisemblance et sa comparaison avec un seuil. Il est évident que lorsque l'on présente les résultats des observations d'un processus aléatoire par un *échantillon discret*, les stratégies mentionnées dans le premier chapitre peuvent être utilisées *directement* pour la vérification des hypothèses statistiques. Ainsi, par exemple, si $W_N[x_1, \ldots, x_N \mid s_0(t)]$ et $W_N[x_1, \ldots, x_N \mid s_1(t)]$ sont les fonctions de vraisemblance de

l'échantillon ($n = 1$), la stratégie (de Bayes, de Neumann-Pearson et parfois du minimax) de vérification des hypothèses sur la valeur moyenne, égale à $s_0(t)$ (hypothèse H_0) ou à $s_1(t)$ (hypothèse H_1), se formule comme suit: on adopte l'hypothèse H_1 si pour les coordonnées observées $x_1, \ldots, x_N$ on a

$$l(x_1, \ldots, x_N) = \frac{W_N[x_1, \ldots, x_N \mid s_1(t)]}{W_N[x_1, \ldots, x_N \mid s_0(t)]} > c, \tag{3.50}$$

où le seuil c dépend du critère de qualité choisi.

Cependant, lors de la vérification des hypothèses statistiques concernant les paramètres des processus aléatoires ayant des densités de probabilité continues de dimensions finies, on peut au lieu des échantillons discrets $x_1, \ldots, x_N$ utiliser toute la réalisation observée $x(t)$ du processus aléatoire $\xi(t)$. Cette possibilité est basée sur un *théorème fondamental* en vertu duquel dans certaines conditions il existe une limite finie en probabilité pour $N \to \infty$ du rapport de vraisemblance $l(x_1, \ldots, x_N)$, que l'ensemble $x_1, \ldots, x_N$ appartienne à la distribution $W_N(x_1, \ldots, x_N \mid H_0)$ ou à $W_N(x_1, \ldots, x_N \mid H_1)$, c'est-à-dire

$$\lim_{N\to\infty} l(x_1, \ldots, x_N) = \lim_{N\to\infty} \frac{W_N(x_1, \ldots, x_N \mid H_1)}{W_N(x_1, \ldots, x_N \mid H_0)} = l[x(t)], \tag{3.51}$$

la limite (3.51) étant différente de zéro. Cette limite est appelée *fonctionnelle du rapport de vraisemblance* de l'échantillon observé. En utilisant la terminologie de Grenander [2], nous appellerons *régulier* le cas où la fonctionnelle du rapport de vraisemblance $l[x(t)] > 0$ existe.

Pour d'autres conditions la fonctionnelle du rapport de vraisemblance augmente indéfiniment ou s'annule, ce qui correspond au cas de la dégénérescence, appelé cas *singulier*.

On peut montrer que dans le cas régulier, pour un temps d'observation quelconque fini T, la stratégie utilisant la fonctionnelle du rapport de vraisemblance conduira obligatoirement à des probabilités non nulles de prise de décisions erronées. Au contraire, dans le cas singulier des décisions correctes sont possibles pour tout $T > 0$.

Comme dans le cas d'un échantillon discret, nous utiliserons non pas la fonctionnelle du rapport de vraisemblance mais son logarithme $\ln l[x(t)]$. Dans le cas régulier $|\ln l[x(t)]|$ est limité, alors que dans le cas singulier il ne l'est pas.

La démonstration générale de ce théorème et de la conséquence mentionnée ci-dessus exige l'emploi d'un appareil mathématique compliqué et sort du cadre de cet ouvrage (cf., par exemple, [2], [4]). Nous donnerons par la suite des applications de ce théorème au cas des processus aléatoires normaux.

3.4.4. Fonctionnelle du rapport de vraisemblance d'un processus aléatoire normal. Soit $\xi(t)$ un processus aléatoire normal de moyenne nulle et de fonction de corrélation $B(t, y)$ connue. Dans le premier cas ce processus est ajouté au processus déterministe $s_0(t)$ (hypothèse H_0) et dans le second, au processus $s_1(t)$ (hypothèse H_1). Ecrivons le logarithme du rapport de vraisemblance pour l'échantillon discret $x_1, \ldots, x_N$ obtenu sur l'intervalle $(-T, T)$ à des intervalles de temps réguliers à partir de la réalisation $x(t)$ de l'un des processus $\xi(t)+s_0(t)$ ou $\xi(t)+s_1(t)$:

$$\ln l(x_1, \ldots, x_N) = -\frac{1}{2D}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{ij}[(x_i-s_{1i})(x_j-s_{1j})-$$
$$-(x_i-s_{0i})(x_j-s_{0j})] = -\frac{1}{2D}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{ij}[(s_{0i}-s_{1i})x_j+$$
$$+(s_{0j}-s_{1j})x_i+(s_{1i}s_{1j}-s_{0i}s_{0j})], \qquad (3.52)$$

où

$$x_k = x(t_k);\quad s_{0k}=s_0(t_k);\quad s_{1k}=s_1(t_k);\quad t_k=\frac{2k-N}{N}T;$$

D et D_{ij} étant le déterminant et le cofacteur de l'élément $B(t_i, t_j)$ de la matrice de corrélation $\mathrm{M} = \| B(t_i, t_j)\|$, on suppose de plus que la matrice M n'est pas dégénérée, c'est-à-dire que $D>0$ et que pour tout N il existe une matrice inverse, dont les éléments sont $\frac{D_{ij}}{D}$.

Nous allons trouver la limite de (3.52) pour $N\to\infty$, c'est-à-dire le logarithme de la fonctionnelle du rapport de vraisemblance. Soit tout d'abord la forme bilinéaire

$$K_N=\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{D_{ij}}{D}x_iy_j. \qquad (3.53)$$

Posons

$$V_j=\sum_{i=1}^{N}\frac{D_{ij}}{D}y_i,\qquad j=1, \ldots, N. \qquad (3.54)$$

On a alors

$$K_N=\sum_{j=1}^{N}V_jx_j. \qquad (3.54')$$

Comme $\frac{D_{ij}}{D}$ sont les éléments de la matrice inverse de la matrice de corrélation, la solution du système d'équations algébriques linéaires (3.54) peut s'écrire comme suit:

$$y_i=\sum_{j=1}^{N}B(t_i, t_j)V_j,\qquad i=1, \ldots, N. \qquad (3.54'')$$

Pour $N \to \infty$ (ou pour max $(t_{k+1} - t_k) \to 0$ sur l'intervalle d'observation) la forme bilinéaire devient une forme intégrale

$$\lim_{N\to\infty} K_N = K = \int_{-T}^{T} V(t)\, x(t)\, dt, \tag{3.55}$$

et le système d'équations linéaires (3.54″), une équation intégrale linéaire non homogène

$$y(t) = \int_{-T}^{T} B(t, u)\, V(u)\, du, \quad |t| \leqslant T, \tag{3.56}$$

à partir de laquelle pour $y(t)$ et $B(t, u)$ donnés on trouve la fonction inconnue $V(t)$ entrant dans l'expression (3.55).

En utilisant (3.55) et (3.56), on peut dans (3.52) passer à la limite pour $N \to \infty$ et obtenir l'expression suivante pour le logarithme de la fonctionnelle du rapport de vraisemblance, vers laquelle le logarithme du rapport de vraisemblance $\ln l(x_1, \ldots, x_N)$ converge en probabilité:

$$\ln l[x(t)] = \int_{-T}^{T} V(t)\, x(t)\, dt + \frac{1}{2}\int_{-T}^{T} [V_0(t)\, s_0(t) - V_1(t)\, s_1(t)]\, dt, \tag{3.57}$$

où $V(t)$, $V_1(t)$, $V_0(t)$ sont les solutions des équations intégrales linéaires non homogènes suivantes (partout $|t| \leqslant T$):

$$\int_{-T}^{T} B(t, u)\, V(u)\, du = s_1(t) - s_0(t), \tag{3.58}$$

$$\int_{-T}^{T} B(t, u)\, V_1(u)\, du = s_1(t), \tag{3.59}$$

$$\int_{-T}^{T} B(t, u)\, V_0(u)\, du = s_0(t). \tag{3.59'}$$

En comparant (3.58) avec (3.59) et (3.59′) on voit que

$$V(t) = V_1(t) - V_0(t). \tag{3.60}$$

Puis, en multipliant les deux membres de (3.59) par $V_0(t)$ et ceux de (3.59′) par $V_1(t)$ et en intégrant sur t de $-T$ à T on obtient

$$\int_{-T}^{T} s_1(t)\, V_0(t)\, dt = \int_{-T}^{T} s_0(t)\, V_1(t)\, dt. \tag{3.60'}$$

En utilisant (3.60) et (3.60′), la formule (3.57) devient

$$\ln l\,[x(t)] = \int_{-T}^{T} V(t)\left[x(t) - \frac{s_0(t)+s_1(t)}{2}\right] dt. \tag{3.61}$$

On peut également obtenir l'expression (3.61) du logarithme de la fonctionnelle du rapport de vraisemblance en passant à la limite pour $N \to \infty$ dans le logarithme du rapport de vraisemblance pour N coordonnées *indépendantes* observées. En utilisant (3.47) on obtient

$$\ln l\,(x_1, \ldots, x_N) = -\frac{1}{2}\sum_{k=1}^{N} [(x_k - b_k)^2 - (x_k - a_k)^2] =$$

$$= \frac{1}{2}\sum_{k=1}^{N} (a_k^2 - b_k^2) + \sum_{k=1}^{N} x_k\,(b_k - a_k), \tag{3.62}$$

où

$$x_k = \sqrt{\lambda_k}\int_{-T}^{T} x(t)\,\varphi_k(t)\,dt; \tag{3.63}$$

$$a_k = \sqrt{\lambda_k}\int_{-T}^{T} s_0(t)\,\varphi_k(t)\,dt; \tag{3.64}$$

$$b_k = \sqrt{\lambda_k}\int_{-T}^{T} s_1(t)\,\varphi_k(t)\,dt; \tag{3.64′}$$

λ_k et $\varphi_k(t)$ étant les nombres caractéristiques et les fonctions propres de l'équation intégrale linéaire homogène (3.23).

Introduisons la notation

$$V_N(t) = \sum_{k=1}^{N} \sqrt{\lambda_k}\,(b_k - a_k)\,\varphi_k(t). \tag{3.65}$$

On peut alors écrire comme suit l'expression (3.62):

$$\ln l\,(x_1, \ldots, x_N) = \int_{-T}^{T} V_N(t)\left[x(t) - \frac{s_0(t)+s_1(t)}{2}\right] dt. \tag{3.66}$$

Notons également qu'en multipliant les deux membres de (3.65) par $B(t, u)$ et en intégrant sur t de $-T$ à T on obtient compte tenu de (3.23)

$$\int_{-T}^{T} B(t, u)\,V_N(t)\,dt = \sum_{k=1}^{N} \frac{b_k - a_k}{\sqrt{\lambda_k}}\,\varphi_k(u). \tag{3.67}$$

Quand

$$\sum_{k=1}^{\infty} (a_k - b_k)^2 < \infty, \tag{3.68}$$

la limite finie

$$\lim_{N\to\infty} V_N(t) = V(t) \tag{3.68'}$$

existe, la fonction $V(t)$ étant donnée par l'équation intégrale non homogène linéaire suivante [cf. (3.67) et (3.31)]

$$\int_{-T}^{T} B(t, u) V(u)\, du = \sum_{k=1}^{\infty} \frac{b_k}{\sqrt{\lambda_k}} \varphi_k(t) - \\ - \sum_{k=1}^{\infty} \frac{a_k}{\sqrt{\lambda_k}} \varphi_k(t) = s_1(t) - s_0(t), \quad |t| \leqslant T,$$

celle-ci coïncidant avec (3.58).

Lorsque la condition (3.68) se trouve remplie, en vertu de (3.68') le logarithme du rapport de vraisemblance (3.66) converge en probabilité vers la fonctionnelle

$$\ln l[x(t)] = \int_{-T}^{T} V(t) \left[x(t) - \frac{s_0(t) + s_1(t)}{2}\right] dt,$$

cette dernière expression coïncidant avec (3.61).

Si la série $\sum_{k=1}^{\infty} (a_k - b_k)^2$ est divergente, on peut montrer (cf., par exemple, [2], § 4.4) que le logarithme du rapport de vraisemblance (3.66) tend en probabilité pour $N \to \infty$ vers $+\infty$, si $x(t)$ appartient au processus $\xi(t) + s_1(t)$ et vers $-\infty$ si $x(t)$ appartient au processus $\xi(t) + s_0(t)$.

Ainsi, si la condition (3.68) se trouve remplie, on a le cas régulier (conformément à la terminologie du § 3.4.3), et si cette condition n'est pas remplie, le cas singulier.

Si l'on observe n réalisations indépendantes $x_1(t), \ldots, x_n(t)$, compte tenu de (3.48) ou de (3.49), on peut généraliser immédiatement (3.61) en écrivant comme suit le logarithme de la fonctionnelle du rapport de vraisemblance:

$$\ln l[x_1(t), \ldots, x_n(t)] = \\ = \int_{-T}^{T} V(t) \left\{\sum_{i=1}^{n} x_i(t) - \frac{n}{2}[s_0(t) + s_1(t)]\right\} dt, \tag{3.69}$$

où $V(t)$ est comme précédemment la solution de l'équation intégrale (3.58).

3.4.5. Généralisation au cas d'un processus aléatoire complexe. Nous allons généraliser les résultats du paragraphe précédent à un processus aléatoire normal complexe $\zeta(t)$ de moyenne nulle et de fonction de corrélation $B_\zeta(t, y)$ connue [cf. (3.34)]. Trouvons le logarithme du rapport de vraisemblance pour un échantillon discret (de taille $2N$) des coordonnées non corrélées

$$z_k = x_k + iy_k, \quad k = 1, \ldots, N, \tag{3.70}$$

obtenues à partir de la réalisation $z(t) = x(t) + iy(t)$ observée sur l'intervalle $(-T, T)$ du processus $\zeta(t) = \xi(t) + i\eta(t)$ (hypothèse H_0) ou du processus $\zeta(t) + s(t) = \xi(t) + a(t) + i[\eta(t) + b(t)]$ où $s(t) = a(t) + ib(t)$ est une fonction complexe déterministe (hypothèse H_1). Il vient de (3.35′)

$$m_1\{x_k x_m\} = m_1\{y_k y_m\} = 0, \ k \neq m,$$

$$m_1\{x_k y_m\} = 0 \text{ pour } k \text{ et } m \text{ quelconques}$$

$$m_1\{x_k^2\} = m_1\{y_k^2\} = \frac{1}{2}.$$

Introduisons de plus les coordonnées du processus déterministe $s_k = a_k + ib_k$.

Comme les variables aléatoires x_k, y_m sont normales, on obtient d'une manière analogue à (3.62)

$$\ln l(x_1, y_1, \ldots, x_N, y_N) = \sum_{k=1}^{N} [x_k^2 + y_k^2 - (x_k - a_k)^2 - (y_k - b_k)^2]$$

ou, sous forme complexe,

$$\ln l(z_1, \ldots, z_N) = \sum_{k=1}^{N} [|z_k|^2 - |z_k - s_k|^2] =$$

$$= 2\sum_{k=1}^{N} \left[\mathrm{Re}\,(s_k \bar{z}_k) - \frac{1}{2}|s_k|^2\right], \tag{3.71}$$

où

$$z_k = \sqrt{\lambda_k} \int_{-T}^{T} z(t)\, \overline{\varphi_k(t)}\, dt; \tag{3.72}$$

$$s_k = \sqrt{\lambda_k} \int_{-T}^{T} s(t)\, \overline{\varphi_k(t)}\, dt, \tag{3.73}$$

et λ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale (3.23) de noyau $B_\zeta(t, u)$.

Introduisons la notation

$$V_N(t) = \sum_{k=1}^{N} \sqrt{\lambda_k}\, s_k \varphi_k(t). \tag{3.74}$$

On peut alors écrire (3.71) comme suit :

$$\ln l(z_1, \ldots, z_N) = 2\,\mathrm{Re}\int_{-T}^{T} V_N(t)\,\overline{\left[z(t) - \frac{s(t)}{2}\right]}\,dt. \qquad (3.75)$$

Si

$$\sum_{k=1}^{\infty} |s_k|^2 < \infty, \qquad (3.76)$$

la limite

$$\lim_{N\to\infty} V_N(t) = V(t)$$

existe. Elle est obtenue à partir de l'équation intégrale linéaire non homogène [comparer avec (3.58)]

$$\int_{-T}^{T} B_\zeta(t, u)\,V(u)\,du = s(t), \quad |t| \leqslant T. \qquad (3.77)$$

Sous forme complexe, l'équation intégrale (3.77) est équivalente à un système de *deux* équations intégrales par rapport à la partie réelle $V_R(t)$ et la partie imaginaire $V_I(t)$ de $V(t)$

$$\int_{-T}^{T} [B_\xi(t, u) + B_\eta(t, u)]\,V_R(u)\,du + \int_{-T}^{T} [B_{\xi\eta}(t, u) - B_{\eta\xi}(t, u)]\,V_I(u)\,du = \mathrm{Re}\,s(t), \qquad (3.77')$$

$$\int_{-T}^{T} [B_\xi(t, u) + B_\eta(t, u)]\,V_I(u)\,du + \int_{-T}^{T} [B_{\eta\xi}(t, u) - B_{\xi\eta}(t, u)]\,V_R(u)\,du = \mathrm{Im}\,s(t). \qquad (3.77'')$$

Lorsque la condition (3.76) se trouve remplie, le logarithme de la fonction de vraisemblance (3.75) converge en probabilité vers la fonctionnelle

$$\ln l[z(t)] = 2\,\mathrm{Re}\int_{-T}^{T} V(t)\,\overline{\left[z(t) - \frac{1}{2}s(t)\right]}\,dt. \qquad (3.78)$$

Lorsque la série (3.76) est divergente, le logarithme du rapport de vraisemblance (3.75) tend vers $+\infty$ pour l'hypothèse H_1, et vers $-\infty$ pour l'hypothèse H_0.

3.4.6. Bruit blanc normal. Cherchons à trouver à titre d'exemple l'expression explicite de la fonctionnelle du rapport de vraisemblance, lorsque $\xi(t)$ est un bruit blanc normal d'intensité N_0 par unité de bande (cf. § 3.4.4). La fonction de corrélation d'un bruit blanc

étant

$$B(t, u) = N_0 \delta (t - u),$$

en vertu de (3.58) on a

$$N_0 V(t) = s_1(t) - s_0(t). \qquad (3.79)$$

On obtient alors à partir de (3.61) l'expression de la fonctionnelle du rapport de vraisemblance pour le cas envisagé

$$\ln l[x(t)] = \frac{1}{N_0} \int_{-T}^{T} [s_1(t) - s_0(t)]\, x(t)\, dt - \\ - \frac{1}{2N_0} \int_{-T}^{T} [s_1^2(t) - s_0^2(t)]\, dt. \qquad (3.80)$$

Pour la fonction de corrélation d'un bruit blanc tous les nombres caractéristiques de l'équation intégrale (3.23) sont $\lambda = \frac{1}{N_0}$ [comparer avec (3.24″)]. La condition (3.68) peut alors s'écrire comme suit:

$$\sum_{k=1}^{\infty} (a_k - b_k)^2 < \infty.$$

Or, la somme des carrés des coefficients $a_k - b_k$ du développement suivant des fonctions orthogonales du processus déterministe $s_0(t) - s_1(t)$ sur l'intervalle $(-T, T)$ est proportionnelle (pour $\lambda = \text{const}$) à la grandeur $\frac{1}{2T} \int_{-T}^{T} [s_0(t) - s_1(t)]^2\, dt$. Ainsi, lorsque la puissance moyenne du processus déterministe se trouve limitée et le processus $\xi(t)$ est un bruit blanc normal, en vertu de (3.68) on *aura toujours le cas régulier.*

3.5. VÉRIFICATION DES HYPOTHÈSES STATISTIQUES CONCERNANT UN PROCESSUS ALÉATOIRE NORMAL

3.5.1. Remarques préliminaires. Nous allons maintenant étudier les stratégies de vérification des hypothèses sur les paramètres des processus aléatoires et leurs estimations, en étendant aux processus aléatoires les méthodes exposées dans les chapitres précédents pour les variables aléatoires. Lorsque les résultats des observations d'un processus aléatoire sont donnés par un échantillon (discret) de taille finie, tout ce qui est utilisé pour obtenir des conclusions statistiques sur les variables aléatoires peut être immédiatement appliqué aux processus aléatoires. C'est pourquoi dans ce qui suit nous nous efforçons à montrer comment sont construites les statistiques déterminant les stratégies et les estimations quand le résultat des obser-

vations est non pas un échantillon discret mais une réalisation continue (ou plusieurs réalisations) du processus aléatoire.

3.5.2. Vérification de l'hypothèse sur la moyenne d'un processus aléatoire normal. Nous allons commencer par le cas de vérification de l'hypothèse simple H_0 selon laquelle la moyenne d'un processus aléatoire normal $\xi(t)$ de fonction de corrélation *) $B(t, u)$ donnée est égale à $s_0(t)$, contre l'alternative simple H_1 que cette moyenne est $s_1(t)$. Supposons tout d'abord que la fonction de corrélation du processus est définie positive ce qui signifie que l'ensemble correspondant des fonctions propres est complet (cf. § 3.3.1).

Comme nous l'avons déjà montré dans le premier chapitre, les différents critères de qualité conduisent à une même procédure de prise de décision: à partir d'un échantillon (discret) de taille donnée on calcule le rapport de vraisemblance et l'on adopte ou l'on rejette l'hypothèse H_0 suivant que ce rapport est supérieur ou non à un certain seuil, dont la grandeur dépend du critère de qualité établi à l'avance et ne dépend pas de la taille de l'échantillon.

Par conséquent, pour vérifier l'hypothèse sur la moyenne d'un processus aléatoire normal, on peut utiliser la stratégie formulée au § 1.3, en remplaçant le rapport de vraisemblance par la fonctionnelle du rapport de vraisemblance (3.61) dont le logarithme est dans le cas régulier borné en valeur absolue. Ainsi, on a la stratégie suivante: si pour la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$ on a

$$\int_{-T}^{T} V(t)\,x(t)\,dt \geqslant \ln c + \frac{1}{2}\int_{-T}^{T} V(t)\,[s_0(t)+s_1(t)]\,dt = k_T, \qquad (3.81)$$

on adopte la décision γ_1 (l'hypothèse H_1 est acceptée), si c'est l'inégalité inverse de (3.81) qui est vérifiée, on adopte la décision γ_0 (l'hypothèse H_1 est rejetée). Dans la formule (3.81) la grandeur c est, suivant le critère adopté, choisie dans la table 1 et la fonction $V(t)$ est la solution de l'équation intégrale linéaire non homogène (3.58).

On peut facilement trouver l'expression des probabilités conditionnelles des erreurs. En effet, à partir de (1.11), (1.12) et (3.81) on a

$$\alpha = P\{\gamma_1 \mid H_0\} = P\left\{\int_{-T}^{T} V(t)\,x(t)\,dt \geqslant k_T \mid s_0(t)\right\}, \qquad (3.82)$$

$$\beta = P\{\gamma_0 \mid H_1\} = P\left\{\int_{-T}^{T} V(t)\,x(t)\,dt < k_T \mid s_1(t)\right\}. \qquad (3.83)$$

Comme $x(t)$ est une réalisation d'un processus aléatoire normal, l'intégrale dans le premier membre de (3.81) est, pour un T donné, une variable aléatoire normale, les moyennes, les seconds moments

*) On suppose ici que $B(t, u) = m_1\{[\xi(t) - m_1\{\xi(t)\}][\xi(u) - m_1\{\xi(u)\}]\}$.

et les variances de cette variable aléatoire étant respectivement

$$m_{10}=m_1\left\{\int_{-T}^{T}V(t)\,x(t)\,dt\mid H_0\right\}=\int_{-T}^{T}V(t)\,s_0(t)\,dt, \tag{3.84}$$

$$m_{11}=m_1\left\{\int_{-T}^{T}V(t)\,x(t)\,dt\mid H_1\right\}=\int_{-T}^{T}V(t)\,s_1(t)\,dt, \tag{3.84'}$$

$$\begin{aligned}m_1\left\{\left(\int_{-T}^{T}V(t)\,x(t)\,dt\right)^2\mid H_0\right\}&=\\&=m_1\left\{\int_{-T}^{T}\int_{-T}^{T}V(t)\,x(t)\,V(u)\,x(u)\,dt\,du\mid H_0\right\}=\\&=\int_{-T}^{T}\int_{-T}^{T}V(t)\,V(u)\,[B(t,\ u)+s_0(t)\,s_0(u)]\,dt\,du=\\&=\int_{-T}^{T}V(t)\,[s_1(t)-s_0(t)]\,dt+\left(\int_{-T}^{T}V(t)s_0(t)\,dt\right)^2,\end{aligned} \tag{3.85}$$

$$\begin{aligned}m_1\left\{\left(\int_{-T}^{T}V(t)\,x(t)\,dt\right)^2\mid H_1\right\}&=\int_{-T}^{T}V(t)\,[s_1(t)-s_0(t)]\,dt+\\&+\left(\int_{-T}^{T}V(t)s_1(t)\,dt\right)^2,\end{aligned} \tag{3.85'}$$

c'est-à-dire

$$\begin{aligned}d_T^2=M_2\left\{\int_{-T}^{T}V(t)\,x(t)\,dt\mid H_0\right\}&=M_2\left\{\int_{-T}^{T}V(t)\,x(t)\,dt\mid H_1\right\}=\\&=\int_{-T}^{T}V(t)\,[s_1(t)-s_0(t)]\,dt=m_{11}-m_{10}.\end{aligned} \tag{3.86}$$

D'où *) [cf. les formules de (3.81) à (3.86)]

$$\begin{aligned}\alpha&=\frac{1}{d_T\sqrt{2\pi}}\int_{\ln c+\frac{m_{10}+m_{11}}{2}}^{\infty}\exp\left[-\frac{(y-m_{10})^2}{2d_T^2}\right]dy=\\&=\int_{\frac{\ln c}{d_T}+\frac{d_T}{2}}^{\infty}\exp\left[-t^2/2\right]dt=1-F\left(\frac{\ln c}{d_T}+\frac{d_T}{2}\right),\end{aligned} \tag{3.87}$$

---

*) Notons également que les grandeurs m_{10}, m_{11}, d_T tout comme $\int_{-T}^{T}V(t)\,x(t)\,dt$ sont *sans dimension*, ce qui découle de l'équation intégrale (3.58).

$$\beta = \frac{1}{d_T \sqrt{2\pi}} \int_{-\infty}^{\ln c + \frac{m_{10}+m_{11}}{2}} \exp\left[-\frac{(y-m_{11})^2}{2d_T^2}\right] dy =$$
$$= F\left(\frac{\ln c}{d_T} - \frac{d_T}{2}\right). \quad (3.88)$$

Si l'on observe non pas une mais n réalisations indépendantes $x_i(t)$, $i = 1, \ldots, n$, en vertu de (3.69) le domaine d'acceptation de l'hypothèse H_1 est donné par l'inégalité

$$\int_{-T}^{T} V(t)\left[\sum_{i=1}^{n} x_i(t)\right] dt \geqslant \ln c + \frac{n}{2} \int_{-T}^{T} V(t)\,[s_0(t) + s_1(t)]\, dt, \quad (3.89)$$

et pour trouver les probabilités conditionnelles d'erreurs on utilise les mêmes formules (3.87) et (3.88) en posant seulement

$$d_T^2 = n \int_{-T}^{T} V(t)\,[s_1(t) - s_0(t)]\, dt. \quad (3.89')$$

Les formules ci-dessus montrent que pour $d_T \to \infty$ la probabilité d'erreurs tend vers zéro (consistance de la stratégie).

Pour distinguer des hypothèses *voisines*, pour un T donné, et obtenir des valeurs acceptables des probabilités d'erreurs, il faut un nombre suffisamment grand de réalisations indépendantes. Supposons, par exemple, que $s_1(t) = a_1 s(t)$ et $s_0(t) = a_0 s(t)$. En introduisant la notation

$$\frac{1}{a_T} = \int_{-T}^{T} V(t)\, s(t)\, dt,$$

on obtient alors à partir de (3.89')

$$d_T^2 = \frac{a_1 - a_0}{a_T} n, \quad (3.89'')$$

par conséquent, pour que la stratégie reste consistante il faut que pour $n \to \infty$ la grandeur $\frac{a_1 - a_0}{a_T}$ décroisse plus lentement que $\frac{1}{n}$.

Quand $\xi(t)$ est un bruit blanc normal de densité spectrale N_0, la fonction $V(t)$ est donnée par (3.79). En portant (3.79) dans (3.81), on obtient la stratégie suivante: on adopte la décision γ_1 si

$$\int_{-T}^{T} [s_1(t) - s_0(t)]\, x(t)\, dt \geqslant N_0 \ln c + \frac{1}{2} \int_{-T}^{T} [s_1^2(t) - s_0^2(t)]\, dt. \quad (3.90)$$

En particulier, lorsque les processus $s_0(t)$ et $s_1(t)$ ont même énergie sur l'intervalle d'observation et $c = 1$ (critère du maximum

de vraisemblance), l'inégalité (3.90) devient

$$\int_{-T}^{T} s_1(t)\,x(t)\,dt \geqslant \int_{-T}^{T} s_0(t)\,x(t)\,dt. \tag{3.90'}$$

La variance de la grandeur $\int_{-T}^{T} V(t)\,x(t)\,dt$, lorsque $x(t)$ correspond à la réalisation de la somme d'un bruit blanc normal et d'un processus déterministe, pour l'une quelconque des hypothèses (H_0 ou H_1), en vertu de (3.86) est égale à

$$d_T^2 = \frac{1}{N_0} \int_{-T}^{T} [s_1(t) - s_0(t)]^2\,dt. \tag{3.91}$$

Montrons maintenant que dans les cas où l'ensemble des fonctions propres *n'est pas complet* on peut indiquer une stratégie assurant des probabilités d'erreurs nulles. Comme nous l'avons déjà mentionné, une telle situation (cas singulier) se présente si l'on renonce à l'hypothèse que la fonction de corrélation $B(t, u)$ est définie positive et si l'on suppose que la forme quadratique (3.21) ou la forme intégrale (3.21') peuvent être nulles.

Par définition, si le système de fonctions orthogonales $\{\varphi_k(t)\}$ n'est pas complet, il existe au moins une fonction $\psi(t)$ orthogonale à toutes les $\varphi_k(t)$. Autrement dit, pour k quelconque on a

$$\int_{-T}^{T} \varphi_k(t)\,\psi(t)\,dt = 0 \tag{3.92}$$

Nous allons utiliser le développement orthogonal du processus aléatoire (3.30). Ecrivons sa réalisation sous la forme

$$x(t) = s_0(t) + \sum_{k=1}^{\infty} x_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \quad |t| \leqslant T,$$

pour l'hypothèse H_0, et sous la forme

$$x(t) = s_1(t) + \sum_{k=1}^{\infty} x_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \quad |t| \leqslant T,$$

pour l'hypothèse H_1.

Il découle de (3.92) que

$$\int_{-T}^{T} x(t)\,\psi(t)\,dt = \begin{cases} c_0, & \text{pour l'hypothèse } H_0, \\ c_1, & \text{pour l'hypothèse } H_1, \end{cases} \tag{3.93}$$

où

$$c_0 = \int_{-T}^{T} s_0(t)\,\psi(t)\,dt; \quad c_1 = \int_{-T}^{T} s_1(t)\,\psi(t)\,dt. \tag{3.93'}$$

Si $c_0 \neq c_1$, on obtient à partir de (3.93) la stratégie absolument correcte (avec une probabilité égale à l'unité)

$$\int_{-T}^{T} x(t)\,\psi(t)\,dt = c_0$$

pour l'hypothèse H_0 et

$$\int_{-T}^{T} x(t)\,\psi(t)\,dt = c_1$$

pour l'hypothèse H_1.

3.5.3. Alternative composite. Nous allons maintenant passer à l'étude d'une alternative composite (comparer avec § 1.4.1). Supposons que l'on vérifie l'hypothèse simple H_0 selon laquelle la moyenne du processus aléatoire normal $\xi(t)$ envisagé au paragraphe précédent (3.5.2) est nulle, contre l'alternative composite H_1 que la moyenne appartient à l'*ensemble* des processus $s_\vartheta(t)$ [ou dans le cas particulier $s(t; \vartheta)$], ϑ pouvant être un nombre réel quelconque (ou se trouver dans certains intervalles de l'axe réel).

Nous allons commencer par la stratégie de Bayes. Dans ce cas il faut se donner les probabilités a priori q et $p = 1 - q$ pour que $m_1\{\xi(t)\} = 0$ et $m_1\{\xi(t)\} = s_\vartheta(t)$ et la densité de probabilité a priori $w_1(\vartheta)$ du paramètre ϑ. En prenant la moyenne du rapport de vraisemblance sur le paramètre aléatoire ϑ pour l'échantillon discret de coordonnées non corrélées $x_1, \ldots, x_N$ on trouve [cf. (1.110') et (3.62)]

$$\Lambda(x_1, \ldots, x_N) = \int_{-\infty}^{\infty} \exp\left\{\sum_{k=1}^{N}\left[x_k b_k(\vartheta) - \frac{1}{2} b_k^2(\vartheta)\right]\right\} w_1(\vartheta)\,d\vartheta, \tag{3.94}$$

où

$$x_k = \sqrt{\lambda_k} \int_{-T}^{T} x(t)\,\varphi_k(t)\,dt; \tag{3.95}$$

$$b_k(\vartheta) = \sqrt{\lambda_k} \int_{-T}^{T} s_\vartheta(t)\,\varphi_k(t)\,dt; \tag{3.95'}$$

λ_k, $\varphi_k(t)$ étant les nombres caractéristiques et les fonctions propres de l'équation intégrale linéaire homogène (3.23).

Comme précédemment, si la somme $\sum_{k=1}^{\infty} \lambda_k b_k^2(\vartheta)$ est finie, la

limite

$$\lim_{N\to\infty} \sum_{k=1}^{N} \sqrt{\lambda_k} b_k(\vartheta) \varphi_k(t) = V(t;\ \vartheta) \tag{3.96}$$

existe, la fonction $V(t;\ \vartheta)$ étant donnée par l'équation intégrale linéaire non homogène

$$\int_{-T}^{T} B(t,\ u) V(u;\ \vartheta)\, du = s_\vartheta(t), \quad |t| \leqslant T. \tag{3.97}$$

Dans ces conditions le rapport moyen de vraisemblance converge en probabilité vers la fonctionnelle

$$\Lambda[x(t)] = \int_{-\infty}^{\infty} \exp\left\{ \int_{-T}^{T} V(t;\ \vartheta)\left[x(t) - \frac{s_\vartheta(t)}{2}\right] dt \right\} w_1(\vartheta)\, d\vartheta. \tag{3.98}$$

En utilisant (3.98) on arrive à la stratégie de Bayes minimisant le risque moyen : adopter la décision γ_1 (l'hypothèse H_0 est rejetée) si pour la réalisation observée on a

$$\int_{-\infty}^{\infty} \exp\left\{ \int_{-T}^{T} V(t;\ \vartheta)\left[x(t) - \frac{s_\vartheta(t)}{2}\right] dt \right\} w_1(\vartheta) \geqslant c = \\ = \frac{q}{p}\,\frac{\Pi_{01} - \Pi_{00}}{\Pi_{10} - \Pi_{11}}, \tag{3.99}$$

et la décision γ_0 (l'hypothèse H_0 est acceptée) si c'est l'inégalité inverse de (3.99) qui est vérifiée [cf. (1.109)].

Dans le cas où $s_\vartheta(t) = a(\vartheta)\, s(t)$ est un processus quasi déterministe d'amplitude aléatoire, la fonctionnelle (3.98) du rapport de vraisemblance peut s'écrire comme suit :

$$\Lambda[x(t)] = \int_{-\infty}^{\infty} \exp\left[a(\vartheta) \int_{-T}^{T} x(t) V(t)\, dt - \\ - \frac{a^2(\vartheta)}{2} \int_{-T}^{T} s(t) V(t)\, dt \right] w_1(\vartheta)\, d\vartheta, \tag{3.99'}$$

où $V(t)$ est la solution de l'équation intégrale

$$\int_{-T}^{T} B(t,\ u) V(u)\, du = s(t), \quad |t| \leqslant T. \tag{3.100}$$

Considérons maintenant le critère de Neumann-Pearson. Des raisonnements analogues permettent d'obtenir pour ϑ quelconque l'expression suivante pour le logarithme de la fonctionnelle du

rapport de vraisemblance:

$$\ln l\,[x(t)] = \int_{-T}^{T} V(t\,;\ \vartheta)\,x(t)\,dt - \frac{1}{2}\int_{-T}^{T} V(t\,;\ \vartheta)\,s_{\vartheta}(t)\,dt. \qquad (3.101)$$

Pour ϑ donné, le domaine critique où l'hypothèse H_0 est rejetée est défini par l'inégalité

$$\int_{-T}^{T} V(t\,;\ \vartheta)\,x(t)\,dt \geqslant K(\vartheta), \qquad (3.102)$$

où $K(\vartheta)$ est obtenu à partir de la valeur de la probabilité d'erreur de première espèce. En vertu de (3.102) pour que le critère uniformément le plus puissant existe (cf. § 1.4.4 et 1.4.7), il faut que

$$s_{\vartheta}(t) = a(\vartheta)\,s(t), \qquad (3.103)$$

l'amplitude $a(\vartheta)$ ne changeant pas de signe. Si $a(\vartheta) > 0$ la stratégie se formule comme suit: l'hypothèse H_0 est rejetée si pour la réalisation observée $x(t)$ on a

$$\int_{-T}^{T} V(t)\,x(t)\,dt \geqslant K, \qquad (3.104)$$

où $V(t)$ est donné par l'équation intégrale (3.100).

Comme dans le cas d'authenticité de l'hypothèse H_0 on a

$$m_1\left\{\int_{-T}^{T} V(t)\,x(t)\,dt\right\} = 0,$$

$$M_2\left\{\int_{-T}^{T} V(t)\,x(t)\,dt\right\} = \int_{-T}^{T} V(t)\,s(t)\,dt = d_T^2,$$

et comme l'intégrale dans (3.104) est répartie normalement, il vient

$$P\left\{\int_{-T}^{T} V(t)\,x(t)\,dt \geqslant K \,|\, H_0\right\} = 1 - F\left(\frac{K}{d_T}\right) \qquad (3.105)$$

et par conséquent, pour une probabilité α d'erreur de première espèce donnée [comparer avec (1.76)] on a

$$K = d_T x_{\alpha}, \qquad (3.105')$$

où x_{α} est un quantile de la loi normale.

Si $a(\vartheta) < 0$, l'inégalité (3.104) est remplacée par

$$\int_{-T}^{T} V(t)\,x(t)\,dt < -K. \qquad (3.106)$$

Ainsi le critère (3.104) est uniformément le plus puissant pour l'alternative composite pour laquelle $a(\vartheta) > 0$, tandis que le critère (3.106) l'est dans le cas où $a(\vartheta) < 0$. Si, au contraire, on prend pour alternative composite toutes les valeurs réelles de l'amplitude $a(\vartheta)$, le critère uniformément le plus puissant n'existe pas.

La fonction de puissance pour le critère (3.104) s'écrit comme suit:

$$1-\beta(\vartheta)=P\left\{\int_{-T}^{T} V(t)\,x(t)\,dt \geqslant K \mid H_1\right\}=$$

$$=1-F\left(\frac{K}{d_T}-a(\vartheta)\,d_T\right). \quad (3.107)$$

Pour $a(\vartheta) > 0$, $1-\beta(\vartheta) > \alpha$ la fonction de puissance croît d'une manière monotone au fur et à mesure de l'augmentation de $a(\vartheta)$. Mais si $a(\vartheta) < 0$, on a $1-\beta(\vartheta) < \alpha$ et la fonction de puissance tend d'une manière monotone vers zéro lorsque $|a(\vartheta)|$ croît indéfiniment. Par conséquent, si le domaine des valeurs possibles de $a(\vartheta)$ s'étend à tous les nombres réels, le critère (3.104) est sans biais.

On peut montrer [2] que, tout comme dans le cas discret étudié au § 1.4.7, le critère uniformément le plus puissant non biaisé qu'on peut construire à partir de la réalisation observée, est donné par la relation suivante:

$$\left|\int_{-T}^{T_1} x(t)\,V(t)\,dt\right| \geqslant K \quad (3.108)$$

3.5.4. Processus à spectres énergétiques en forme de fraction rationnelle. L'exposé des paragraphes précédents montre que la difficulté essentielle de recherche des critères de vérification de l'hypothèse sur la moyenne d'un processus aléatoire normal est liée à ce que la relation du critère en question avec la fonction de corrélation donnée $B(t, u)$ du processus n'est pas explicite, mais exprimée par l'intermédiaire de la solution $V(t)$ de l'équation intégrale linéaire non homogène (3.58). La solution de cette équation est dans le cas général inconnue, parfois on peut l'obtenir au prix de calculs forts compliqués, et évidemment il reste toujours la possibilité d'utiliser les méthodes numériques. Seul fait exception le cas où le processus normal se présente sous la forme de la somme d'une fonction déterministe et d'un processus aléatoire normal stationnaire $\xi(t)$ de moyenne nulle et dont le spectre énergétique $F_\xi(\omega)$ est une fraction rationnelle de la forme

$$F_\xi(\omega)=\frac{Q(\omega)}{P(\omega)},$$

où Q et P sont des polynômes de degrés m et n respectivement $(m < n)$. Le processus $\xi(t)$ est alors la *composante* d'un processus

markovien normal *multidimensionnel* (cf. § 9.5.1, tome I). Dans le cas mentionné l'équation intégrale (3.58) se réduit à une équation différentielle ordinaire à coefficients constants et sa solution peut être écrite sous forme explicite (cf., par exemple, [3], annexe II).

Considérons le problème de la vérification de l'hypothèse simple H_0 suivant laquelle la réalisation observée appartient au processus $\xi(t) + s_0(t)$ contre l'alternative simple H_1 qu'elle appartient à $\xi(t) + s_1(t)$, à condition que $\xi(t)$ soit un processus markovien normal stationnaire. Dans ce cas, comme nous l'avons déjà noté au § 4.5.4 du premier tome, le spectre énergétique du processus est

$$F_\xi(\omega) = \frac{4\mu\sigma^2}{\mu^2+\omega^2},$$

et sa fonction de corrélation

$$B(\tau) = \sigma^2 e^{-\mu|\tau|}, \quad \mu > 0. \tag{3.109}$$

La solution de l'équation intégrale (3.58) dont le noyau est du type (3.109) s'écrit comme suit:

$$V(t) = \frac{1}{2\mu\sigma^2}\{[\mu s_\Delta(-T) - s'_\Delta(-T)]\,\delta(t+T) + \\ + [\mu s_\Delta(T) - s'_\Delta(T)]\,\delta(t-T) + \mu^2 s_\Delta(t) - s''_\Delta(t)\}, \quad |t| \leqslant T, \tag{3.110}$$

où

$$s_\Delta(t) = s_1(t) - s_0(t). \tag{3.110'}$$

En portant (3.110') dans (3.81) et en introduisant la notation

$$s_\Sigma(t) = s_1(t) + s_0(t), \tag{3.110''}$$

pour la somme des fonctions déterministes, on peut alors formuler la stratégie comme suit: adopter la décision γ_1 (acceptation de l'hypothèse H_1) si pour la réalisation $x(t)$ du processus, observée sur l'intervalle $(-T, T)$, on a

$$\left[s_\Delta(-T) - \frac{s'_\Delta(-T)}{\mu}\right]x(-T) + \left[s_\Delta(T) - \frac{s'_\Delta(T)}{\mu}\right]x(T) + \\ + \mu\int_{-T}^{T} x(t)\left[s_\Delta(t) - \frac{s''_\Delta(t)}{\mu^2}\right]dt \geqslant 2\sigma^2 \ln c + \frac{1}{2}\left\{\left[s_\Delta(-T) - \right.\right. \\ \left. - \frac{s'_\Delta(-T)}{\mu}\right]s_\Sigma(-T) + \left[s_\Delta(T) - \frac{s'_\Delta(T)}{\mu}\right]s_\Sigma(T) + \\ \left. + \mu\int_{-T}^{T} s_\Sigma(t)\left[s_\Delta(t) - \frac{s''_\Delta(t)}{\mu^2}\right]dt\right\}. \tag{3.111}$$

Les probabilités conditionnelles d'erreurs de première et de seconde espèces sont données par la grandeur [cf. (3.87) et (3.88)]

$$d_T^2 = \frac{1}{2\sigma^2}\left\{\left[s_\Delta(-T) - \frac{s'_\Delta(-T)}{\mu}\right]s_\Delta(-T) + \right.$$

$$\left. + \left[s_\Delta(T) - \frac{s'_\Delta(T)}{\mu}\right]s_\Delta(T) + \mu\int_{-T}^{T}\left[s_\Delta(t) - \frac{s''_\Delta(t)}{\mu^2}\right]s_\Delta(t)\,dt\right\} =$$

$$= \frac{1}{2\sigma^2}\left\{s_\Delta^2(-T) + s_\Delta^2(T) + \mu\int_{-T}^{T}\left(s_\Delta^2(t) + \frac{1}{\mu^2}[s'_\Delta(t)]^2\right)dt\right\}. \quad (3.112)$$

Si pour $\mu T \gg 1$ on a $\frac{s_\Delta(|T|)}{\sigma} < 1$ et $\frac{s'_\Delta(|T|)}{\mu\sigma} < 1$, on obtient à partir de (3.111) et (3.112) les expressions asymptotiques suivantes, vraies pour un temps d'observation suffisamment long:

$$\mu\int_{-T}^{T} x(t)\left[s_\Delta(t) - \frac{s''_\Delta(t)}{\mu^2}\right]dt \geqslant 2\sigma^2 \ln c +$$

$$+ \frac{\mu}{2}\int_{-T}^{T} s_\Sigma(t)\left[s_\Delta(t) - \frac{s''_\Delta(t)}{\mu^2}\right]dt, \quad (3.113)$$

$$d_T^2 \sim \frac{\mu}{2\sigma^2}\int_{-T}^{T}\left\{s_\Delta^2(t) + \frac{1}{\mu^2}[s'_\Delta(t)]^2\right\}dt. \quad (3.114)$$

Dans le cas où, pour t quelconque, les fonctions $s_0(t)$ et $s_1(t)$ sont constantes, c'est-à-dire si

$$s_0(t) \equiv a_0, \quad s_1(t) \equiv a_1, \quad (3.115)$$

le problème envisagé se réduit à la vérification de l'hypothèse sur la grandeur de la moyenne d'un processus aléatoire normal dont la fonction de corrélation est de la forme (3.109). Les dérivées des fonctions $s_0(t)$ et $s_1(t)$ sont alors identiquement nulles et à partir de (3.111) on arrive à la conclusion suivante: la moyenne est égale à a_1 si

$$x(-T) + x(T) + \mu\int_{-T}^{T} x(t)\,dt \geqslant \frac{2\sigma^2}{a_1 - a_0}\ln c + (a_1 + a_0)(1 + \mu T), \quad (3.116)$$

avec

$$d_T^2 = \frac{(a_1 - a_0)^2}{\sigma^2}(1 + \mu T). \quad (3.117)$$

Si $\mu T \gg 1$ on a les relations asymptotiques

$$\frac{x(-T)+x(T)}{2\mu T}+\frac{1}{2T}\int_{-T}^{T}(t)\,dt > \frac{a_1+a_0}{2}, \qquad (3.118)$$

la grandeur d_T^2 augmentant proportionnellement à μT

$$d_T^2 \sim \left(\frac{a_1-a_0}{\sigma}\right)^2 \mu T. \qquad (3.118')$$

3.5.5. Vérification d'une hypothèse sur la fonction de corrélation. Soit à vérifier l'hypothèse H_0 selon laquelle la fonction de corrélation d'un processus aléatoire normal $\xi(t)$ de moyenne nulle *) est égale à $B_0(t, y)$, contre l'alternative H_1 que cette fonction de corrélation est égale à $B_1(t, y)$. Si l'on prend pour coordonnées du processus $\xi(t)$ les variables aléatoires [cf. (3.25)]

$$\xi_k = \sqrt{\lambda_k}\int_{-T}^{T}\xi(t)\,\varphi_k(t)\,dt, \qquad (3.119)$$

où λ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale

$$\varphi(t)=\lambda\int_{-T}^{T}B_0(t, u)\,\varphi(u)\,du, \quad |t|\leqslant T,$$

ces coordonnées ne sont pas corrélées si c'est l'hypothèse H_0 qui est vraie et *corrélées* si c'est l'hypothèse H_1 qui est vraie.

Cependant, on peut choisir les coordonnées du processus de telle sorte qu'elles ne soient pas corrélées (et, par conséquent, indépendantes, la répartition étant normale) tant pour l'hypothèse H_0 que pour l'hypothèse H_1, avec cette seule différence que pour l'une des hypothèses les variances de toutes les coordonnées sont égales à l'unité, et pour l'autre, elles diffèrent pour différentes coordonnées. Ce cas est analogue à un résultat bien connu en algèbre supérieure en vertu duquel on peut, par une seule transformation linéaire, ramener l'une forme quadratique à une forme normale (c'est-à-dire à la somme des carrés des variables) et une autre, à une forme canonique **).

Supposons que les coordonnées du processus $\xi(t)$ soient données par (3.119), λ_k et $\varphi_k(t)$ étant les nombres caractéristiques et les fonctions propres (non normées) de l'équation intégrale

$$\int_{-T}^{T}[B_1(t, u)-\lambda B_0(t, u)]\,\varphi(u)\,du=0, \quad |t|\leqslant T, \qquad (3.120)$$

---

*) Il est facile de généraliser les résultats de ce paragraphe en supposant que la moyenne du processus est différente de zéro [12]. Le problème 1.6 est l'analogue discret de ce problème.

**) Voir, par exemple, A. Michina et I. Proskouriakov. *Algèbre supérieure* (en russe). Physmathguiz, 1962, pp. 162 à 164.

la normalisation des fonctions propres étant effectuée par rapport à $B_0(t, y)$, c'est-à-dire que l'on a

$$\lambda_k \int_{-T}^{T}\int_{-T}^{T} B_0(t, u)\varphi_k(t)\varphi_m(u)\,dt\,du = \begin{cases} 1, & k = m, \\ 0, & k \neq m. \end{cases} \tag{3.120'}$$

On a alors

$$m_1\{\xi_k\xi_m \mid H_0\} = \sqrt{\lambda_k\lambda_m}\int_{-T}^{T}\int_{-T}^{T} B_0(t, u)\varphi_k(t)\varphi_m(u)\,dt\,du = \begin{cases} 1, & k = m, \\ 0, & k \neq m, \end{cases} \tag{3.121}$$

$$m_1\{\xi_k\xi_m \mid H_1\} = \sqrt{\lambda_k\lambda_m}\int_{-T}^{T}\int_{-T}^{T} B_1(t, u)\varphi_k(t)\varphi_m(u)\,dt\,du =$$
$$= \lambda_m\sqrt{\lambda_k\lambda_m}\int_{-T}^{T}\int_{-T}^{T} B_0(t, u)\varphi_k(t)\varphi_m(u)\,dt\,du = \begin{cases} \lambda_k, & k = m, \\ 0, & k \neq m. \end{cases} \tag{3.121'}$$

En utilisant (3.121) et (3.121') il est facile d'écrire le logarithme du rapport de vraisemblance pour N coordonnées indépendantes $x_1, \ldots, x_N$ observées d'une réalisation d'un processus aléatoire normal stationnaire, soit:

$$\ln l(x_1, \ldots, x_N) = -\frac{1}{2}\sum_{k=1}^{N}\ln\lambda_k - \frac{1}{2}\sum_{k=1}^{N}\left(\frac{x_k^2}{\lambda_k} - x_k^2\right) =$$
$$= \frac{1}{2}\sum_{k=1}^{N}\frac{\lambda_k - 1}{\lambda_k}x_k^2 - \frac{1}{2}\sum_{k=1}^{N}\ln\lambda_k, \tag{3.122}$$

où en vertu de (3.119)

$$x_k = \sqrt{\lambda_k}\int_{-T}^{T} x(t)\varphi_k(t)\,dt, \tag{3.122'}$$

λ_k et $\varphi_k(t)$ étant les nombres caractéristiques et les fonctions propres de l'équation intégrale (3.120), et $x(t)$ la réalisation d'un processus aléatoire normal observée sur l'intervalle $(-T, T)$.

En portant (3.122') dans (3.122) et en passant à la limite pour $N \to \infty$ on obtient l'expression suivante pour le logarithme de la fonctionnelle du rapport de vraisemblance:

$$\ln l[x(t)] = \frac{1}{2}\sum_{k=1}^{\infty}(\lambda_k - 1)\left[\int_{-T}^{T} x(t)\varphi_k(t)\,dt\right]^2 - \frac{1}{2}\sum_{k=1}^{\infty}\ln\lambda_k. \tag{3.123}$$

On peut maintenant formuler la stratégie: si pour la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$ on a

$$\sum_{k=1}^{\infty}(\lambda_k-1)\left[\int_{-T}^{T} x(t)\,\varphi_k(t)\,dt\right]^2 \geqslant 2\ln c+\sum_{k=1}^{\infty}\ln\lambda_k. \qquad (3.124)$$

on adopte la décision γ_1 (l'hypothèse H_1 est vraie), et si c'est l'inégalité inverse de (3.124) qui est vérifiée on adopte la décision γ_0 (l'hypothèse H_0 est vraie). Comme d'habitude, le seuil c est choisi conformément au critère de qualité adopté (voir table 1, page 35).

Pour trouver les probabilités d'erreurs, il y a lieu avant tout de calculer la répartition du logarithme de la fonctionnelle du rapport de vraisemblance (3.123). Bien que $\ln l\,[x(t)]$ soit une somme infinie de variables aléatoires indépendantes $\frac{\lambda_k-1}{2\lambda_k}x_k^2-\frac{1}{2}\ln\lambda_k$, la répartition de cette somme *n'est pas normale* car la condition rendant applicable le théorème limite de Liapounov ne se trouve pas remplie [cf. (3.144), tome I]. Pour trouver cette répartition nous allons utiliser la méthode des fonctions caractéristiques. Ainsi, comme x_k est une variable aléatoire normale, les fonctions caractéristiques unidimensionnelles de la k-ième composante de la somme pour les hypothèses H_0 et H_1 sont respectivement [cf. (9.16), tome I]

$$\Theta_{1k}(v\,|\,H_0)=\frac{1}{\sqrt{1-iv\left(1-\frac{1}{\lambda_k}\right)}}\,e^{-\frac{iv}{2}\ln\lambda_k}, \qquad (3.125)$$

$$\Theta_{1k}(v\,|\,H_1)=\frac{1}{\sqrt{1-iv\,(\lambda_k-1)}}\,e^{-\frac{iv}{2}\ln\lambda_k}. \qquad (3.125')$$

Les variables aléatoires x_k étant indépendantes, les fonctions caractéristiques unidimensionnelles du logarithme de la fonctionnelle du rapport de vraisemblance sont données par le produit des fonctions caractéristiques (3.125) et (3.125')

$$\Theta_1(v\,|\,H_0)=\left[\prod_{k=1}^{\infty}\left(1-\frac{iv}{1+v_k}\right)\right]^{-\frac{1}{2}}\times$$
$$\times\exp\left\{-\frac{iv}{2}\ln\prod_{k=1}^{\infty}\left(1+\frac{1}{v_k}\right)\right\}, \qquad (3.126)$$

$$\Theta_1(v\,|\,H_1)=\left[\prod_{k=1}^{\infty}\left(1-\frac{iv}{v_k}\right)\right]^{-\frac{1}{2}}\times$$
$$\times\exp\left\{-\frac{iv}{2}\ln\prod_{k=1}^{\infty}\left(1+\frac{1}{v_k}\right)\right\}. \qquad (3.126')$$

avec

$$v_k = \frac{1}{\lambda_k - 1}. \qquad (3.126'')$$

On peut maintenant, par transformation de Fourier inverse, trouver les densités de probabilité et les fonctions de répartition du logarithme de la fonctionnelle de rapport de vraisemblance, bien que les calculs soient en général très ardus.

Arrêtons-nous encore sur le cas où l'on connaît le coefficient de corrélation $R(\tau)$ d'un processus aléatoire normal stationnaire et où l'on a à vérifier seulement l'hypothèse H_0 que la variance est égale à σ_0^2 contre l'alternative H_1 qu'elle est égale à σ_1^2. Dans le cas envisagé, on a $B_0(\tau) = \sigma_0^2 R(\tau)$, $B_1(\tau) = \sigma_1^2 R(\tau)$. Introduisons la notation $\left(\frac{\sigma_1}{\sigma_0}\right)^2 = \mu^2$ et supposons que $\mu \neq 1$. On obtient alors à partir de (3.120)

$$(\mu^2 - \lambda) \int_{-T}^{T} R(t-u)\,\varphi(u)\,du = 0, \qquad (3.127)$$

donc les nombres caractéristiques λ_k de l'équation intégrale sont constants et égaux à μ^2. En vertu de (3.122) le logarithme du rapport de vraisemblance peut s'écrire comme suit:

$$\ln l(x_1, \ldots, x_N) = \frac{1}{2}\left(1 - \frac{1}{\mu^2}\right) \sum_{k=1}^{N} x_k^2 - \frac{N}{2} \ln \mu^2,$$

et, par conséquent,

$$\lim_{N\to\infty} \frac{1}{N} \ln l(x_1, \ldots, x_N) = \frac{1}{2}\left(1 - \frac{1}{\mu^2}\right) \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x_k^2 - \ln \mu =$$

$$= \begin{cases} \frac{1}{2}\left(1 - \frac{1}{\mu^2}\right) - \ln \mu < 0, & \text{pour l'hypothèse } H_0, \\ \frac{1}{2}(\mu^2 - 1) - \ln \mu > 0, & \text{pour l'hypothèse } H_1, \end{cases} \qquad (3.128)$$

car la suite des variables aléatoires $z_N = \frac{1}{N} \sum_{k=1}^{N} x_k^2$ converge en probabilité vers l'unité *) pour l'hypothèse H_0 et vers μ^2 pour

*) Notons que le théorème de Liapounov est applicable pour la somme z_N, alors pour $N \to \infty$ on a

$$W_1(z_N) \sim \frac{1}{\sqrt{2\pi M_2\{z_N\}}} \exp\left\{-\frac{1}{2M_2\{z_N\}} [z_N - m_1\{z_N\}]^2\right\},$$

avec $m_1\{z_N \mid H_0\} = 1$, $m_1\{z_N \mid H_1\} = \mu^2$, $M_2\{z_N \mid H_0\} = \frac{2}{N}$, $M_2\{z_N \mid H_1\} = \frac{2\mu^2}{N}$. A la limite on a $\lim_{N\to\infty} W_1(z_N) = \delta(z - m_1\{z\})$.

l'hypothèse H_1. En vertu de (3.128) on a pour l'hypothèse H_0

$$\lim_{N\to\infty} \ln l(x_1, \ldots, x_N) = -\infty, \tag{3.129}$$

et pour l'hypothèse H_1

$$\lim_{N\to\infty} \ln l(x_1, \ldots, x_N) = \infty. \tag{3.129'}$$

Ainsi, le cas envisagé est singulier, on peut donc, avec une probabilité égale à l'unité, choisir pour un intervalle de temps d'observation arbitraire (aussi petit que l'on veut) la stratégie de vérification de l'hypothèse. Cette stratégie découle immédiatement de (3.128) si l'on substitue aux coordonnées x_k leur expression (3.122') en fonction de la réalisation observée. Si pour la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$ on a

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \lambda_k \left(\int_{-T}^{T} x(t)\,\varphi_k(t)\,dt \right)^2 = 1, \tag{3.130}$$

on accepte l'hypothèse H_0 (la variance du processus est égale à σ_0^2) et si la limite dans (3.130) est égale à μ^2, on accepte l'hypothèse H_1 (la variance du processus est égale à σ_1^2).

3.5.6. Cas où l'hypothèse vérifiée se rapporte à un bruit blanc. Considérons un cas particulier du problème général exposé au paragraphe précédent. Supposons que l'on vérifie l'hypothèse suivant laquelle $\xi(t)$ est un bruit blanc normal de densité spectrale N_0 contre l'alternative que c'est un processus aléatoire normal et que sa fonction de corrélation est égale à $B(t, y) + N_0\delta(t - y)$. Nous allons montrer que dans ce cas on peut remplacer dans (3.124) la sommation par l'intégration des réalisations pondérées par une fonction de pondération dépendant de $B(t, y)$.

Dans ce cas l'équation intégrale (3.120) s'écrit comme suit:

$$N_0(\lambda - 1)\,\varphi(t) = \int_{-T}^{T} B(t, y)\,\varphi(y)\,dy, \quad |t| \leqslant T, \tag{3.131}$$

la normalisation étant obtenue par division par $\sqrt{\lambda_k N_0}$ de chacune des fonctions propres $\varphi_k(t)$ [cf. (3.120')]. Introduisons la notation

$$h(t, y) = \sum_{k=1}^{\infty} \frac{\lambda_k - 1}{\lambda_k}\,\varphi_k(t)\,\varphi_k(y), \tag{3.132}$$

où $\varphi_k(t)$ sont des fonctions orthogonales et *normées*. Le premier terme dans (3.122) s'écrira alors

$$\sum_{k=1}^{\infty} \frac{\lambda_k - 1}{\lambda_k} x_k^2 = \int_{-T}^{T}\int_{-T}^{T} x(t)\,x(y) \sum_{k=1}^{\infty} \frac{\lambda_k - 1}{\lambda_k N_0} \times$$

$$\times\, \varphi_k(t)\,\varphi_k(y)\,dt\,dy = \frac{1}{N_0} \int_{-T}^{T}\int_{-T}^{T} h(t, y)\,x(t)\,x(y)\,dt\,dy. \tag{3.133}$$

Puis à partir de (3.132) on obtient

$$\int_{-T}^{T} [B(t, y)+N_0\delta(t-y)]\, h(v, y)\, dy = \sum_{k=1}^{\infty} \frac{\lambda_k-1}{\lambda_k} \varphi_k(v) \times$$
$$\times \int_{-T}^{T} [B(t, y)+N_0\delta(t-y)]\, \varphi_k(y)\, dy = \sum_{k=1}^{\infty} N_0(\lambda_k-1)\,\varphi_k(v) \times$$
$$\times \varphi_k(t) = B(t, v).$$

Par conséquent, la fonction de pondération $h(t, u)$ satisfait à l'équation intégrale

$$\int_{-T}^{T} B(t, y)\, h(v, y)\, dy + N_0 h(v, t) = B(t, v), \quad |t| \leqslant T, \ |v| \leqslant T. \tag{3.134}$$

L'équation intégrale (3.134) est un cas particulier (pour $u = 1$) de l'équation plus générale

$$u \int_{-T}^{T} B(t, y)\, h(v, y, u)\, dy + N_0 h(v, t, u) = B(t, v), \quad |t| \leqslant T, \quad |v| \leqslant T \tag{3.135}$$

par rapport à la fonction $h(v, t, u)$ appelée *résolvante*. Pour la résolvante on a le développement

$$h(t, y, u) = \sum_{k=1}^{\infty} \frac{\lambda_k-1}{1+(\lambda_k-1)u} \varphi_k(t)\, \varphi_k(y), \tag{3.136}$$

donnant (3.132) pour $u = 1$. En vertu de (3.136) on a

$$\int_{-T}^{T} h(t, t, u)\, dt = \sum_{k=1}^{\infty} \frac{\lambda_k-1}{1+(\lambda_k-1)u}, \tag{3.136'}$$

d'où par intégration sur la variable u de zéro à l'unité on obtient

$$\int_0^1 \int_{-T}^{T} h(t, t, u)\, dt\, du = \int_0^1 \left(\sum_{k=1}^{\infty} \frac{\lambda_k-1}{1+(\lambda_k-1)u} \right) du =$$
$$= \sum_{k=1}^{\infty} \int_0^{\lambda_k-1} \frac{dx}{1+x} = \sum_{k+1}^{\infty} \ln \lambda_k.$$

Ainsi, le second terme dans le second membre de (3.123) est dans le cas envisagé égal à

$$\sum_{k=1}^{\infty} \ln \lambda_k = \int_0^1 \int_{-T}^{T} h(t, t, u)\, dt\, du. \tag{3.137}$$

En portant (3.133) et (3.137) dans (3.124) on obtient la stratégie suivante de vérification de l'hypothèse H_0 selon laquelle la réalisation observée appartient à un bruit blanc: l'hypothèse H_0 est rejetée si

$$\frac{1}{N_0}\int_{-T}^{T}\int_{-T}^{T} h(t, y)\, x(t)\, x(y)\, dt\, dy \geqslant 2\ln c + \int_{0}^{1}\int_{-T}^{T} h(t, t, u)\, dt\, du, \quad (3.138)$$

où $h(t, y)$ et $h(t, y, u)$ sont respectivement les solutions des équations intégrales (3.134) et (3.135).

Pour trouver dans ce cas les fonctions caractéristiques du logarithme du rapport de vraisemblance nous allons utiliser les formules (3.126) et (3.126′). A cet effet introduisons tout d'abord la fonction $\ln D(z)$ égale à l'intégrale de (3.136′) sur u prise de zéro à z, c'est-à-dire

$$\ln D(z) = \int_{0}^{z}\int_{-T}^{T} h(t, t, u)\, dt\, du =$$

$$= \sum_{k=1}^{\infty} \ln\left(1+\frac{z}{v_k}\right) = \ln \prod_{k=1}^{\infty}\left(1+\frac{z}{v_k}\right) \quad (3.139)$$

ou

$$D(z) = \exp\left\{\int_{0}^{z}\int_{-T}^{T} h(t, t, u)\, dt\, du\right\} = \prod_{k=1}^{\infty}\left(1+\frac{z}{v_k}\right), \quad (3.139')$$

ici v_k est donné par (3.126″).

En utilisant (3.139) et (3.139′) on obtient les expressions cherchées pour les fonctions caractéristiques [cf. (3.126), (3.126′)]

$$\Theta_1(v \mid H_0) = \sqrt{\frac{D(1)}{D(1-iv)}} \exp\left\{-\frac{iv}{2}\ln D(1)\right\}, \quad (3.140)$$

$$\Theta_1(v \mid H_1) = \frac{1}{\sqrt{D(-iv)}} \exp\left\{-\frac{iv}{2}\ln D(1)\right\}. \quad (3.140')$$

Les probabilités conditionnelles d'erreurs de première et de seconde espèce relatives à l'utilisation de (3.138) sont

$$\alpha = \frac{\sqrt{D(1)}}{2\pi}\int_{\ln c}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{D(1-iv)}} \exp\{-iv[x+\ln\sqrt{D(1)}]\}\, dv\, dx, \quad (3.141)$$

$$\beta = \frac{1}{2\pi}\int_{-\infty}^{\ln c}\int_{-\infty}^{\infty} \frac{1}{\sqrt{D(-iv)}} \exp\{-iv[x+\ln\sqrt{D(1)}]\}\, dv\, dx. \quad (3.141')$$

3.5.7. Généralisation aux processus à spectres énergétiques en forme de fraction rationnelle. On peut appliquer les résultats du paragraphe précédent au cas où l'hypothèse vérifiée et l'alternative se rapportent à des processus aléatoires normaux dont les spectres énergétiques s'expriment par des fractions rationnelles (cf. § 3.5.4). Pour que la relation (3.133) reste vraie il faut que la fonction de pondération $h(t, y)$ satisfasse à l'équation intégrale bidimensionnelle

$$\frac{1}{N_0}\int_{-T}^{T}\int_{-T}^{T} h(s, y)\, B_0(t-s)\, B_1(y-v)\, ds\, dy = B_1(t-v)-B_0(t-v), \quad |t|\leqslant T, \ |v|\leqslant T, \tag{3.142}$$

où N_0 est un facteur constant (dimensionné). Pour $B_1(t-v) = B(t-v) + N_0\delta(t-v)$, $B_0(t-v) = N_0\delta(t-v)$ l'équation (3.142) devient (3.134) (sous la condition que le processus soit stationnaire et le spectre énergétique $F_1(\omega)$ rationnel).

En introduisant la transformée de Fourier bidimensionnelle de la fonction $\frac{1}{N_0}h(s, y)$

$$\frac{1}{N_0}h(s, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(\omega, \lambda)\, e^{i(\lambda s-\omega v)}\, d\lambda\, d\omega,$$

on peut écrire l'équation intégrale (3.142) comme suit (cf. [13]):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(\omega, \lambda)\, F_0(\omega)\, F_1(\lambda)\, e^{-i(\lambda t-\omega v)}\, d\omega\, d\lambda = B_1(t-v)-B_0(t-v), \quad |t|\leqslant T, \ |v|\leqslant T, \tag{3.142'}$$

où $F_0(\omega)$ et $F_1(\omega)$ sont les spectres énergétiques des deux processus aléatoires stationnaires normaux, chacun d'eux pouvant s'écrire sous la forme

$$F(\omega) = \left|\frac{G(i\omega)}{V(i\omega)}\right|^2, \tag{3.143}$$

toutes les racines des polynômes $G(i\omega)$ et $V(i\omega)$ ayant des parties réelles négatives et le degré du polynôme du numérateur n'étant pas supérieur à celui du dénominateur.

Désignons par

$$L(\lambda, v) = \int_{-\infty}^{\infty} H(\omega, \lambda)\, F_0(\omega)\, e^{i\omega v}\, d\omega, \tag{3.144}$$

l'intégrale dans (3.142') sur la variable ω. Nous avons pour l'équation intégrale (3.142')

$$\int_{-\infty}^{\infty} L(\lambda, v)\, F_1(\omega)\, e^{-i\lambda t}\, d\lambda = B_1(t-v)-B_0(t-v), \quad |t|\leqslant T, \ |v|\leqslant T. \tag{3.145}$$

Les équations (3.144) et (3.145) sont deux équations intégrales linéaires non homogènes du même type, se réduisant chacune à une équation différentielle linéaire (cf. [13]). On résout d'abord l'équation (3.145) en prenant la variable v pour paramètre, puis on trouve la solution de l'équation (3.144) par rapport à $H(\omega, \lambda)$ en prenant pour paramètre la variable λ.

La stratégie générale exprimée par (3.124) s'écrit ici d'une manière analogue à (3.138) : l'hypothèse H_0 est rejetée si

$$\frac{1}{N_0}\int_{-T}^{T}\int_{-T}^{T} h(t, y)\,x(t)\,x(y)\,dt\,dy \geqslant 2\ln c + \sum_{k=1}^{\infty} \ln \lambda_k, \qquad (3.146)$$

le premier terme du second membre de (3.146) dépend comme toujours du critère de qualité, la détermination du second terme conduit à des calculs supplémentaires assez compliqués. Comme il est démontré dans [13], les grandeurs $\delta_k = \frac{\lambda_k}{1-\lambda_k}$ sont les nombres caractéristiques de l'équation différentielle suivante (écrite sous la forme opérationnelle) :

$$G_0\left(\frac{d}{dt}\right) G_0\left(-\frac{d}{dt}\right) V_1\left(\frac{d}{dt}\right) V_1\left(-\frac{d}{dt}\right) y(t) =$$
$$= \left(1+\frac{1}{\delta}\right) G_1\left(\frac{d}{dt}\right) G_1\left(-\frac{d}{dt}\right) V_0\left(\frac{d}{dt}\right) V_0\left(-\frac{d}{dt}\right) y(t)$$

pour les conditions aux limites correspondantes. La solution générale de cette équation est de la forme

$$y(t) = \sum_{k=1}^{2N} c_k \exp\{tz_k(\delta)\},$$

où N est le degré du polynôme $G_0(i\omega)\,V_1(i\omega)$, et $z_k(\delta)$ sont les racines du polynôme caractéristique correspondant à l'équation différentielle ci-dessus. Les constantes c_k sont données par le système d'équations algébriques linéaires obtenues pour les conditions aux limites. Si $D(\delta)$ est le déterminant de ce système d'équations et

$$E(\delta) = \frac{D(\delta)}{\prod_{k>j}[z_k(\delta) - z_j(\delta)]}.$$

le second terme du deuxième membre de l'inégalité (3.146) peut s'écrire comme suit [13] :

$$\sum_{k=1}^{\infty} \ln \lambda_k = \frac{E(-1)}{E(0)}. \qquad (3.146')$$

Notons que dans le cas envisagé, lorsque le spectre énergétique d'un processus aléatoire normal stationnaire (au sens général) est

une fraction rationnelle, fonction de la fréquence, est formulée la condition nécessaire et suffisante de singularité [1], [20] applicable lors de la vérification de l'hypothèse H_0 d'après laquelle la réalisation observée appartient à un processus de spectre énergétique $F_0(\omega)$ contre l'alternative H_1 qu'elle appartient à un processus de spectre énergétique $F_1(\omega)$. Cette condition s'écrit

$$\lim_{\omega\to\infty} \frac{F_1(\omega)}{F_0(\omega)} \neq 1. \qquad (3.147)$$

La règle singulière (3.130) *) formulée plus haut correspond au cas particulier de (3.147), où $\frac{F_1(\omega)}{F_0(\omega)} \equiv \mu^2 \neq 1$. Si pour $\omega \to \infty$ la limite du rapport de deux spectres énergétiques en forme de fractions rationnelles est égale à l'unité, on aura toujours le cas régulier auquel correspondent les probabilités non nulles de décisions erronées. La condition suffisante de singularité est également l'existence d'un intervalle fini de fréquences sur lequel au moins l'un des spectres énergétiques $F_1(\omega)$ ou $F_0(\omega)$ soit identiquement nul. Ainsi, le modèle mathématique d'un processus aléatoire à spectre énergétique limité (cf. § 3.2.3) conduit à une singularité.

Notons enfin qu'on aura toujours le cas régulier si (pour n'importe laquelle des deux hypothèses) le processus normal contient une composante additive sous forme d'un *bruit blanc* d'intensité égale, car la condition (3.147) de décision correcte lors de la vérification des hypothèses est basée sur la différence des parties H.F. du spectre énergétique. Comme le bruit blanc est présent dans tout dispositif réel (par exemple, bruits thermiques), le fait de l'ajouter élimine le *paradoxe de singularité* et rapproche le modèle mathématique du processus physique étudié (cf. également [21]).

3.6. ESTIMATIONS DES CARACTÉRISTIQUES D'UN PROCESSUS ALÉATOIRE

3.6.1. Estimation des paramètres de la répartition d'un processus aléatoire d'après sa réalisation. Dans ce paragraphe nous allons étudier un autre type de probabilités statistiques concernant les caractéristiques de la répartition d'un processus aléatoire d'après une réalisation de ce processus observée sur un intervalle de temps fini. Il s'agit de l'estimation des paramètres inconnus des fonctions uni- et multidimensionnelles de la répartition du processus aléatoire. Tout comme au § 3.5 nous avons étendu aux processus aléatoires la théorie de la vérification des hypothèses statistiques d'après des échantillons discrets de taille finie, extraits d'une répartition

*) Notons que la règle singulière (3.130) est vraie non seulement pour un spectre énergétique décrivable par une fraction rationnelle mais également pour tout spectre énergétique satisfaisant à la condition $F_1(\omega) = \mu^2 F_0(\omega)$.

des variables aléatoires, nous allons donner ici une généralisation de la théorie des estimations des paramètres des lois de répartition des variables aléatoires (cf. § 2.2).

Au contraire des estimations des paramètres de la densité de probabilité d'une variable aléatoire, qui sont des fonctions de plusieurs variables des valeurs échantillonnées [cf. (2.14)], les estimations (statistiques) des paramètres $\vartheta_1, \ldots, \vartheta_m$ de la densité de probabilité d'un processus aléatoire sont des *fonctionnelles* de sa réalisation $x(t)$:

$$\hat{\vartheta}_i = g_i\,[x(t)], \quad i = 1, \ldots, m. \tag{3.148}$$

La consistance et l'absence de biais des estimations (3.148) se définissent, comme aux § 2.2.2 et 2.2.3, sous cette réserve que le centrage s'effectue non pas dans l'espace à dimensions finies des échantillons, mais dans l'espace *fonctionnel* des réalisations *).

Les définitions ultérieures et les résultats de la théorie des estimations des paramètres des variables aléatoires sont liés à la fonction de vraisemblance qui est la densité de probabilité d'un point de l'espace des échantillons [cf. (3.43)]. Comme nous l'avons déjà souligné à la fin du § 3.4.1, la fonction de vraisemblance n'est pas bornée lorsque les dimensions de l'espace des échantillons augmentent indéfiniment. Ceci paraît être un obstacle à la généralisation formelle des résultats de la théorie des estimations, donnée au second chapitre, au cas des processus aléatoires. Pour surmonter cet obstacle, on introduit, pour construire la théorie des estimations des paramètres des processus aléatoires, au lieu de la fonctionnelle de vraisemblance qui n'existe pas, la *fonctionnelle du rapport de vraisemblance* **) qui, dans le cas régulier, est positive et limitée [cf. (3.51)].

Ainsi, l'information de Fisher sur le paramètre estimé ϑ contenue dans la réalisation $x(t)$ du processus aléatoire est donnée par la relation

$$I(\vartheta) = m_1\left\{\left(\frac{\partial}{\partial\vartheta}\ln l\,[x(t)\,|\,\vartheta]\right)^2\right\}, \tag{3.149}$$

où m_1 est la moyenne prise sur l'ensemble des réalisations, et $l\,[x(t)\,|\,\vartheta]$ la fonctionnelle du rapport de vraisemblance pour une valeur donnée du paramètre. Le minimum de la variance de l'estimation est donné par l'inégalité généralisée de Rao-Cramer (cf. § 2.2.5)

$$M_2\{\hat{\vartheta}\} \geqslant \frac{\left[1 + \frac{db(\vartheta)}{d\vartheta}\right]^2}{I(\vartheta)}, \tag{3.150}$$

*) Sur l'intégration dans l'espace fonctionnel voir, par exemple, I. Helfand et A. Iaglom, *Intégration dans les espaces fonctionnels*. Uspekhi math. nauk, 1956, 1 (67).

**) Pour plus de détails cf. [2] et [4].

où le *biais* est égal à

$$b(\vartheta) = m_1 \{g[x(t)]\} - \vartheta. \tag{3.151}$$

L'estimation pour laquelle il y a égalité dans (3.150) est appelée *efficace*.

L'estimation du maximum de vraisemblance du paramètre ϑ est donnée par l'équation suivante [cf. (2.54)]:

$$\frac{\partial \ln l[x(t) \mid \vartheta]}{\partial \vartheta} = 0. \tag{3.152}$$

Si l'estimation efficace existe, elle correspond au maximum de vraisemblance, l'affirmation inverse n'étant en général pas vraie.

Les estimations de Bayes peuvent être généralisées de la même manière. Si le paramètre ϑ est aléatoire et si l'on connaît sa densité de probabilité a priori $w_1(\vartheta)$, on peut calculer la densité de probabilité a posteriori du paramètre ϑ d'après la réalisation $x(t)$ du processus aléatoire

$$W_1[\vartheta \mid x(t)] = \frac{w_1(\vartheta)\, l[x(t) \mid \vartheta]}{\int\limits_{-\infty}^{\infty} w_1(\vartheta)\, l[x(t) \mid \vartheta] d\vartheta} \tag{3.153}$$

et puis calculer le risque moyen (risque de Bayes)

$$R = m_1 \left\{ \int\limits_{-\infty}^{\infty} \Pi(g[x(t)], \vartheta) \quad W_1[\vartheta \mid x(t)]\, d\vartheta \right\}, \tag{3.154}$$

où $\Pi(\hat{\vartheta}, \vartheta)$ est la fonction de pertes (cf. § 2.3.4)].

En introduisant la notation

$$J[g \mid x(t)] = \int\limits_{-\infty}^{\infty} \Pi(g, \vartheta)\, W_1[\vartheta \mid x(t)]\, d\vartheta, \tag{3.155}$$

on trouve l'estimation de Bayes $\hat{\vartheta} = g[x(t)]$ du paramètre ϑ à partir de la condition

$$\frac{\partial J}{\partial g} = 0. \tag{3.156}$$

Les notions et les résultats ci-dessus peuvent tout naturellement être généralisés au cas des estimations simultanées d'un ensemble de paramètres $\vartheta_1, \ldots, \vartheta_m$. Nous donnerons ces généralisations plus bas, dans des problèmes concrets.

3.6.2. Estimations du maximum de vraisemblance des paramètres d'une composante déterministe. Soit $x(t)$ une réalisation observée sur l'intervalle $(-T, T)$ d'un processus *somme* du processus aléatoire $\xi(t)$ de moyenne nulle et de fonction de corrélation $B(t, y)$ donnée et du processus déterministe $s(t; \vartheta_1, \ldots, \vartheta_m)$, fonction des paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$. Cherchons à *estimer* les paramè-

tres inconnus $\vartheta_1, \ldots, \vartheta_m$ de la composante déterministe à partir de la réalisation observée $x(t)$. A cet effet nous allons utiliser l'expression du logarithme de la fonctionnelle du rapport de vraisemblance (3.61) en posant $s_1(t) = s(t; \vartheta_1, \ldots, \vartheta_m)$, $s_0(t) \equiv 0$:

$$\ln l[x(t) \mid \vartheta_1, \ldots, \vartheta_m] = \int_{-T}^{T} V(t; \vartheta_1, \ldots, \vartheta_m) \left[x(t) - \frac{1}{2} s(t; \vartheta_1, \ldots, \vartheta_m) \right] dt, \quad (3.157)$$

où $V(t)$ est la solution de l'équation intégrale linéaire non homogène

$$\int_{-T}^{T} B(t, u) V(u; \vartheta_1, \ldots, \vartheta_m) \, du = s(t; \vartheta_1, \ldots, \vartheta_m), \quad |t| \leqslant T. \quad (3.158)$$

En prenant pour critère de qualité des estimations $\hat{\vartheta}_1, \ldots, \hat{\vartheta}_m$ des paramètres inconnus le maximum de la fonctionnelle du rapport de vraisemblance, on trouve à partir de (3.157) un système d'équations auquel ces estimations doivent satisfaire:

$$\frac{\partial}{\partial \vartheta_i} \ln [x(t) \mid \vartheta_1, \ldots, \vartheta_m] = 0, \quad i = 1, \ldots, m,$$

ou

$$\int_{-T}^{T} \frac{\partial}{\partial \vartheta_i} V(t; \vartheta_1, \ldots, \vartheta_m) \left[x(t) - \frac{1}{2} s(t; \vartheta_1, \ldots, \vartheta_m) \right] dt - \\ - \frac{1}{2} \int_{-T}^{T} V(t; \vartheta_1, \ldots, \vartheta_m) \frac{\partial}{\partial \vartheta_i} s(t; \vartheta_1, \ldots, \vartheta_m) \, dt = 0. \quad (3.159)$$

Mais en vertu de (3.158) on a

$$\int_{-T}^{T} V(t; \vartheta_1, \ldots, \vartheta_m) \frac{\partial}{\partial \vartheta_i} s(t; \vartheta_1, \ldots, \vartheta_m) \, dt = \int_{-T}^{T} V(t; \vartheta_1, \ldots \\ \ldots, \vartheta_m) \left[\int_{-T}^{T} B(t, u) \frac{\partial}{\partial \vartheta_i} V(u; \vartheta_1, \ldots, \vartheta_m) \, du \right] dt = \\ = \int_{-T}^{T} \frac{\partial}{\partial \vartheta_i} V(u; \vartheta_1, \ldots, \vartheta_m) \left[\int_{-T}^{T} B(t, u) V(t; \vartheta_1, \ldots, \vartheta_m) \, dt \right] du = \\ = \int_{-T}^{T} \frac{\partial}{\partial \vartheta_i} V(u; \vartheta_1, \ldots, \vartheta_m) s(u; \vartheta_1, \ldots, \vartheta_m) \, du. \quad (3.160)$$

Portons (3.160) dans (3.159), il vient finalement

$$\int_{-T}^{T} \frac{\partial}{\partial \vartheta_i} V(t; \vartheta_1, \ldots, \vartheta_m) [x(t) - s(t; \vartheta_1, \ldots, \vartheta_m)] \, dt = 0, \\ i = 1, \ldots, m. \quad (3.161)$$

En résolvant le système d'équations (3.161) par rapport aux inconnues $\vartheta_1, \ldots, \vartheta_m$, on peut trouver les estimations conditionnelles de ces paramètres $\bar{\vartheta}_i = g_i [x(t)]$, $i = 1, \ldots, m$ satisfaisant au critère adopté du maximum de vraisemblance.

On peut facilement généraliser le système d'équations (3.161) du maximum de vraisemblance au cas de la représentation complexe des processus. En utilisant (3.78) on obtient

$$\int_{-T}^{T} \frac{\partial}{\partial \vartheta_i} V(t; \vartheta_1, \ldots, \vartheta_m) \overline{[z(t) - s(t; \vartheta_1, \ldots, \vartheta_m)]}\, dt +$$

$$+ \int_{-T}^{T} \frac{\partial}{\partial \vartheta_i} \overline{V(t; \vartheta_1, \ldots, \vartheta_m)} [z(t) - s(t; \vartheta_1, \ldots, \vartheta_m)]\, dt = 0, \quad (3.161')$$

où $V(t; \vartheta_1, \ldots, \vartheta_m)$ est la solution d'une équation intégrale du type (3.77).

Si le processus $\xi(t)$ est un bruit blanc de densité spectrale N_0, on a en vertu de (3.158)

$$V(t; \vartheta_1, \ldots, \vartheta_m) = \frac{1}{N_0} s(t; \vartheta_1, \ldots, \vartheta_m) \quad (3.162)$$

et le système (3.161) se simplifie

$$\int_{-T}^{T} \frac{\partial}{\partial \vartheta_i} s(t; \vartheta_1, \ldots, \vartheta_m) [x(t) - s(t; \vartheta_1, \ldots, \vartheta_m)]\, dt = 0,$$

$$i = 1, \ldots, m. \quad (3.163)$$

Dans le cas particulier d'un seul paramètre inconnu ϑ, pour $m = 1$ la formule (3.161) devient

$$\int_{-T}^{T} \frac{\partial V(t; \vartheta)}{\partial \vartheta} [x(t) - s(t; \vartheta)]\, dt = 0, \quad (3.164)$$

et pour un bruit blanc la formule (3.163) s'écrit

$$\int_{-T}^{T} \frac{\partial s(t; \vartheta)}{\partial \vartheta} [x(t) - s(t; \vartheta)]\, dt = 0. \quad (3.165)$$

3.6.3. Estimation de l'amplitude. A titre d'exemple nous allons étudier l'estimation de l'amplitude du processus déterministe $as(t)$. L'équation du maximum de vraisemblance (3.164) donnant cette estimation s'écrit comme suit:

$$\int_{-T}^{T} V(t) [x(t) - as(t)]\, dt = 0.$$

d'où

$$\hat{a} = \frac{\int_{-T}^{T} V(t)\, x(t)\, dt}{\int_{-T}^{T} V(t)\, s(t)\, dt}, \tag{3.166}$$

$V(t)$ étant en vertu de (3.158) la solution de l'équation intégrale linéaire

$$\int_{-T}^{T} B(t, u)\, V(u)\, du = s(t), \quad |t| \leqslant T. \tag{3.167}$$

L'estimation du maximum de vraisemblance de l'amplitude d'un processus déterministe noyé dans un bruit blanc normal est obtenue en portant (3.162) dans (3.166) pour $m = 1$, il vient:

$$\hat{a} = \frac{\int_{-T}^{T} s(t)\, x(t)\, dt}{\int_{-T}^{T} s^2(t)\, dt}. \tag{3.168}$$

Notons que pour $s(t) \equiv 1$ il découle de (3.168)

$$\hat{a} = \frac{1}{2T} \int_{-T}^{T} x(t)\, dt, \tag{3.169}$$

c'est-à-dire qu'on obtient l'estimation d'une moyenne constante a en prenant la moyenne dans le temps de la réalisation observée sur l'intervalle $(-T, T)$. Par analogie avec le cas discret, on appelle parfois l'estimation (3.169) *moyenne arithmétique*.

Introduisons la notation

$$h(t) = \frac{V(t)}{\int_{-T}^{T} V(t)\, s(t)\, dt}. \tag{3.170}$$

L'estimation (3.166) s'écrira alors comme suit:

$$\hat{a} = \int_{-T}^{T} h(t)\, x(t)\, dt. \tag{3.171}$$

Par conséquent, l'estimation du maximum de vraisemblance de l'amplitude de la composante déterministe s'obtient par intégration pondérée, sur l'intervalle d'observation, de la réalisation du processus aléatoire, le poids dépendant de la forme de la composante déterministe et de la fonction de corrélation du processus. De telles estimations sont dites *linéaires*.

Nous allons étudier maintenant les propriétés de l'estimation (3.171). Avant tout il est évident que cette estimation est absolument correcte (non biaisée). En effet, ayant fixé T, on a de (3.170) et (3.171)

$$m_1\{\hat{a}\} = \int_{-T}^{T} h(t)\, m_1\{x(t)\}\, dt = a \int_{-T}^{T} h(t)\, s(t)\, dt = a, \qquad (3.172)$$

car $m_1\{x(t)\} = as(t)$.

La variance de l'estimation $\hat{a}$ est égale à

$$M_2\{\hat{a}\} = m_1\{(\hat{a} - a)^2\} =$$

$$= \int_{-T}^{T} \int_{-T}^{T} h(t)\, h(u)\, m_1\{[x(t) - as(t)]\,[x(u) - as(u)]\}\, dt\, du =$$

$$= \int_{-T}^{T} \int_{-T}^{T} h(t)\, h(u)\, B(t, u)\, dt\, du \qquad (3.173)$$

ou, compte tenu de (3.170) et (3.167),

$$M_2\{\hat{a}\} = \frac{\int_{-T}^{T} \int_{-T}^{T} V(t)\, V(u)\, B(t, u)\, dt\, du}{\left(\int_{-T}^{T} V(t)\, s(t)\, dt\right)^2} = \frac{1}{\int_{-T}^{T} V(t)\, s(t)\, dt}. \qquad (3.174)$$

Cependant, le dénominateur du second membre de (3.174) coïncide avec la moyenne du carré de la dérivée par rapport au paramètre estimé du logarithme de la fonctionnelle du rapport de vraisemblance. En effet, l'information de Fisher est en vertu de (3.149) égale à

$$I(a) = m_1\left\{\left(\frac{\partial}{\partial a} \ln l\,[x(t), a]\right)^2\right\} =$$

$$= m_1\left\{\int_{-T}^{T} \int_{-T}^{T} V(t)\, V(u)\, [x(t) - as(t)]\,[x(u) - as(u)]\, dt\, du\right\} =$$

$$= \int_{-T}^{T} \int_{-T}^{T} V(t)\, V(u)\, B(t, u)\, dt\, du = \int_{-T}^{T} V(t)\, s(t)\, dt. \qquad (3.175)$$

Par conséquent, l'inégalité de Rao-Cramer [cf. (3.150)] permet de conclure que l'estimation linéaire non biaisée (3.171) a une variance minimale dans la classe de toutes les estimations possibles de l'amplitude de la composante déterministe, c'est-à-dire que c'est une *estimation efficace* de ce paramètre.

Comme l'estimation (3.171) du maximum de vraisemblance est une fonctionnelle linéaire de la réalisation d'un processus aléatoire

normal, elle est une variable aléatoire de moyenne a (car elle n'est pas biaisée) et de variance $M_2\{\hat{a}\}$ donnée par (3.174). Ceci permet de donner une *estimation d'intervalle* pour l'amplitude [cf. (2.52)]

$$P\{a(1-\varepsilon)<\hat{a}<a(1+\varepsilon)\}=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\frac{a\varepsilon}{\sqrt{M_2}}}^{\frac{a\varepsilon}{\sqrt{M_2}}} e^{-\frac{u^2}{2}}\,du=2F\left(\frac{a\varepsilon}{\sqrt{M_2}}\right)-1=\gamma, \quad (3.176)$$

par conséquent, la longueur de l'intervalle de confiance est [cf. (3.174)]

$$2\varepsilon a=2\left[\int_{-T}^{T} V(t)\,s(t)\,dt\right]^{-\frac{1}{2}} x_{\frac{1-\gamma}{2}}, \quad (3.176')$$

où $x_{\frac{1-\gamma}{2}}$ est le quantile de la loi normale calculé à partir du coefficient de confiance γ.

Dans le cas d'un bruit blanc on a $N_0V(t) = s(t)$ et la variance (minimale) de l'estimation $\hat{a}$ est donnée par la formule

$$M_2\{\hat{a}\}=\frac{N_0}{\int_{-T}^{T} s^2(t)\,dt}. \quad (3.177)$$

Notons que, bien que l'estimation (3.171) soit l'estimation du maximum de vraisemblance de l'amplitude d'un processus déterministe additionné à un processus aléatoire *normal*, cette estimation possède également certaines propriétés importantes lorsque les processus *ne sont pas normaux*. Tout d'abord il est évident que cette estimation est toujours non biaisée. De plus, on peut montrer que, dans la classe des estimations linéaires, sa variance est minimale. En effet, supposons que l'estimation de l'amplitude soit égale à

$$\hat{a}=\int_{-T}^{T} g(t)\,x(t)\,dt, \quad (3.178)$$

en l'absence de biais on a alors

$$\int_{-T}^{T} g(t)\,s(t)\,dt=1. \quad (3.178')$$

D'une manière analogue à (3.173)

$$M_2\{\hat{a}\}=\int_{-T}^{T}\int_{-T}^{T} g(t)\,g(u)\,B(t,u)\,dt\,du. \quad (3.179)$$

Nous allons trouver maintenant l'infimum de la variance pour différentes *) fonctions $g(t)$ sous la condition (3.178′) qui, compte tenu de (3.167), peut s'écrire comme suit:

$$\int_{-T}^{T}\int_{-T}^{T} g(t)\,V(u)\,B(t, u)\,dt\,du = 1.$$

En utilisant l'inégalité de Bouniakovski-Schwartz (voir, par exemple, la note à la page 76, tome I) on obtient

$$1 = \left(\int_{-T}^{T}\int_{-T}^{T} g(t)\,V(t)\,B(t, u)\,dt\,du\right)^2 \leqslant$$

$$\leqslant \int_{-T}^{T}\int_{-T}^{T} V(t)\,V(u)\,B(t, u)\,dt\,du \int_{-T}^{T}\int_{-T}^{T} g(t)\,g(u)\,B(t, u)\,dt\,du,$$

d'où [cf. (3.179)]

$$M_2\{\hat{a}\} \geqslant \frac{1}{\int_{-T}^{T}\int_{-T}^{T} V(t)\,V(u)\,B(t, u)\,dt\,du}$$

ou [cf. (3.174)]

$$M_2\{\hat{a}\} \geqslant \frac{1}{\int_{-T}^{T} V(t)\,s(t)\,dt}. \tag{3.180}$$

Ainsi, l'infimum des variances des différentes estimations linéaires de l'amplitude a du processus déterministe $as(t)$ additionné à un processus aléatoire arbitraire de fonction de corrélation $B(t, u)$ est égal à $\left(\int_{-T}^{T} V(t)\,s(t)\,dt\right)^{-1}$ où $V(t)$ est la solution de l'équation intégrale (3.167). Cette limite correspond à $g(t) \equiv h(t)$ [cf. (3.170)]. Dans tout ce qui vient d'être dit, on supposait que

$$\int_{-T}^{T} V(t)\,s(t)\,dt = \int_{-T}^{T}\int_{-T}^{T} V(t)\,V(u)\,B(t, u)\,dt\,du > 0, \tag{3.181}$$

c'est-à-dire que la fonction de corrélation est le noyau positif d'une forme quadratique définie positive. Comme nous l'avons noté plus haut, cette condition correspond au cas régulier. Au contraire, si

$$\int_{-T}^{T} V(t)\,s(t)\,dt = 0,$$

*) On suppose évidemment que la classe des fonctions $g(t)$ est limitée non seulement par la condition (3.178′) mais également par l'exigence que $M_2\{\hat{a}\}$ soit borné, c'est-à-dire que l'intégrale stochastique (1.178) converge en moyenne quadratique.

on a le cas singulier. La présence d'un bruit blanc additif suffit pour que l'hypothèse (3.181) se trouve vérifiée. En effet, pour

$$B(t, u) = B_1(t, u) + N_0\delta(t - u)$$

on a

$$\int_{-T}^{T} V(t) s(t)\,dt = \int_{-T}^{T}\int_{-T}^{T} V(t) V(u) B_1(t, u)\,dt\,du + $$

$$+ N_0 \int_{-T}^{T} V^2(t)\,dt > 0, \qquad (3.181')$$

ceci parce que le premier terme n'est pas négatif et le second est positif.

Ainsi, la présence d'un bruit blanc additif exclut immédiatement l'apparition d'une singularité.

3.6.4. Estimations de Bayes. En généralisant la méthode exposée au § 2.4.5, on peut facilement trouver les estimations de Bayes des paramètres aléatoires $\vartheta_1, \ldots, \vartheta_m$ de la composante quasi déterministe $s(t; \vartheta_1, \ldots, \vartheta_m)$ d'après la réalisation observée $x(t)$.

Les estimations de Bayes $\hat{\vartheta}_1 = g_1[x(t)], \ldots, \hat{\vartheta}_m = g_m[x(t)]$ des paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$ minimisant le risque moyen

$$R = m_1 \{J[x(t) \mid g_1, \ldots, g_m]\}, \qquad (3.182)$$

sont données par le système d'équations [cf. (2.118)]

$$\frac{\partial J}{\partial g_i} = 0, \qquad i = 1, \ldots, m, \qquad (3.183)$$

où

$$J[g_1, \ldots, g_m \mid x(t)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \Pi(\hat{\vartheta}_1, \ldots, \hat{\vartheta}_m, \vartheta_1, \ldots, \vartheta_m) \times$$

$$\times W_m[\vartheta_1, \ldots, \vartheta_m \mid x(t)]\,d\vartheta_1 \ldots d\vartheta_m \qquad (3.184)$$

et

$$W_m[\vartheta_1, \ldots, \vartheta_m \mid x(t)] =$$

$$= \frac{w_m(\vartheta_1, \ldots, \vartheta_m)\, l[x(t) \mid \vartheta_1, \ldots, \vartheta_m]}{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w_m(\vartheta_1, \ldots, \vartheta_m)\, l[x(t) \mid \vartheta_1, \ldots, \vartheta_m]\,d\vartheta_1 \ldots d\vartheta_m}. \qquad (3.185)$$

Pour une fonction de pertes quadratique on a [cf. (2.119)]

$$\hat{\vartheta}_i = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \vartheta_i W_m[\vartheta_1, \ldots, \vartheta_m \mid x(t)]\,d\vartheta_1 \ldots d\vartheta_m. \qquad (3.186)$$

Pour les limitations mentionnées dans le second chapitre, l'estimation de Bayes est la même pour d'autres fonctions de pertes (symétriques).

Revenant à l'exemple étudié ci-dessus, calculons l'estimation de Bayes de l'amplitude a du processus quasi déterministe $as(t)$, en supposant tout d'abord que la fonction de pertes Π est quadratique et que l'amplitude est une variable aléatoire normale de moyenne a_0 et de variance σ_0^2 données.

En vertu de (3.157) et (3.185) la densité de probabilité a posteriori du paramètre a, si l'on observe un segment de la réalisation $x(t)$ sur l'intervalle $(-T, T)$, est égale à

$$W_1[a \mid x(t)] = \frac{w_1(a)\, l[x(t) \mid a]}{\int\limits_{-\infty}^{\infty} w_1(a)\, l[x(t) \mid a]\, da} =$$

$$= \frac{1}{\sigma_0 \sqrt{2\pi}} e^{\frac{-(a-a_0)^2}{2\sigma_0^2}} \exp\left\{a \int\limits_{-T}^{T} V(t)\left[x(t) - \frac{a}{2} s(t)\right] dt\right\} \times$$

$$\times \left\{\frac{1}{\sigma_0 \sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{\frac{-(a-a_0)^2}{2\sigma_0^2}} \exp\left[a \int\limits_{-T}^{T} V(t)\left[x(t) - \frac{a}{2} s(t)\right] dt\right] da\right\}^{-1},$$

où $V(t)$ est la solution de l'équation intégrale (3.167). En calculant l'intégrale se trouvant dans l'accolade on obtient

$$W_1[a \mid x(t)] = \frac{(1+\sigma_0^2 s_T)^{1/2}}{\sigma_0 \sqrt{2\pi}} \times$$

$$\times \exp\left\{-\frac{1+\sigma_0^2 s_T}{2\sigma_0^2}\left(a - \frac{a_0 + \sigma_0^2 x_T}{1+\sigma_0^2 s_T}\right)^2\right\}, \qquad (3.187)$$

où

$$s_T = \int\limits_{-T}^{T} V(t)\, s(t)\, dt, \qquad (3.188)$$

$$x_T = \int\limits_{-T}^{T} V(t)\, x(t)\, dt. \qquad (3.189)$$

Pour une fonction de pertes quadratique *) le minimum de $J[x(t) \mid g]$ correspond à l'estimation coïncidant avec la moyenne conditionnelle du paramètre estimé a (cf. § 2.3.6). Il vient immédiatement de (3.187) que l'estimation de Bayes de l'amplitude de la composante quasi

*) En vertu du § 2.3.8, les résultats qui suivent sont également vrais pour toute fonction de pertes symétrique.

déterministe est

$$\hat{a}=\frac{a_0+\sigma_0^2\int_{-T}^{T}V(t)\,x(t)\,dt}{1+\sigma_0^2\int_{-T}^{T}V(t)\,s(t)\,dt}=\frac{a_0+\sigma_0^2 s_T\frac{x_T}{s_T}}{1+\sigma_0^2 s_T} \tag{3.190}$$

ou

$$\hat{a}=a_0+\frac{\sigma_0^2}{1+\sigma_0^2 s_T}\int_{-T}^{T}V(t)\,[x(t)-a_0 s(t)]\,dt=$$

$$=a_0+\frac{\sigma_0^2}{1+\sigma_0^2 s_T}(x_T-a_0 s_T). \tag{3.190'}$$

En comparant (3.190) et (3.166) on voit que l'estimation de Bayes est la moyenne pondérée des deux grandeurs: de l'estimation du maximum de vraisemblance et de la moyenne a priori a_0, de plus le rapport du poids attribué à la première grandeur au poids de la seconde est égal à

$$\mu=\sigma_0^2 s_T=\frac{\sigma_0^2}{\left(\int_{-T}^{T}V(t)\,s(t)\,dt\right)^{-1}}, \tag{3.191}$$

c'est-à-dire au rapport de la variance de la répartition a priori du paramètre à la variance de l'estimation du maximum de vraisemblance.

Si le rapport mentionné augmente indéfiniment (c'est-à-dire si la variance de l'estimation du maximum de vraisemblance est inférieure à celle de la répartition a priori) on a

$$\hat{a}\sim\frac{x_T}{s_T}, \tag{3.191'}$$

ce qui veut dire que l'estimation de Bayes tend vers l'estimation du maximum de vraisemblance. Si au contraire, la variance de la répartition a priori est nettement inférieure à celle du maximum de vraisemblance on a

$$\hat{a}\sim a_0, \tag{3.192}$$

c'est-à-dire que la réalisation observée n'influe pas sur l'estimation qui est supposée égale à la moyenne a priori du paramètre estimé.

Lorsque la composante aléatoire du processus est un bruit blanc, en vertu de (3.167) on a $N_0V(t)=s(t)$ et (3.190) donne

$$\hat{a}=\frac{a_0+\frac{\sigma_0^2}{N_0}\int_{-T}^{T}s(t)\,x(t)\,dt}{1+\frac{\sigma_0^2}{N_0}\int_{-T}^{T}s^2(t)\,dt} \tag{3.192'}$$

ou

$$\hat{a} = \frac{a_0 + \mu \hat{a}_{\text{mv}}}{1 + \mu}, \tag{3.192''}$$

où $\hat{a}_{\text{mv}}$ est l'estimation du maximum de vraisemblance (3.168). La grandeur

$$\mu = \frac{\sigma_0^2}{N_0} \int_{-T}^{T} s^2(t)\, dt$$

est égale au rapport de la variance de la répartition a priori et à celle de l'estimation du maximum de vraisemblance [cf. (3.177)].

Nous allons attirer attention sur certaines propriétés de l'estimation de Bayes (3.190) de l'amplitude de la composante quasi déterministe. Avant tout on remarque que cette estimation s'obtient par transformation *linéaire* de la réalisation observée $x(t)$ et, par conséquent, est, tout comme le processus initial, normalement répartie (comme l'est aussi l'estimation du maximum de vraisemblance). Nous allons calculer la moyenne et la variance de l'estimation de Bayes $\hat{a}$. Calculons d'abord la moyenne et la variance conditionnelles ayant fixé a, puis trouvons les grandeurs inconditionnelles prenant la moyenne sur a. En vertu de (3.190') on a

$$m_1\{\hat{a} \mid a\} = a_0 + (a - a_0)\frac{\sigma_0^2 s_T}{1 + \sigma_0^2 s_T},$$

$$M_2\{\hat{a} \mid a\} = \frac{\sigma_0^4 s_T}{(1 + \sigma_0^2 s_T)^2}\,[1 + (a - a_0)^2 s_T],$$

et comme $m_1\{a\} = a_0$, $m_1\{(a - a_0)^2\} = \sigma_0^2$, en prenant la moyenne sur a, on obtient

$$m_1\{\hat{a}\} = a_0, \tag{3.193}$$

$$M_2\{\hat{a}\} = \frac{\sigma_0^4 s_T}{1 + \sigma_0^2 s_T} = \frac{\mu}{1 + \mu}\sigma_0^2. \tag{3.193'}$$

Pour $\mu \to 0$ la variance de l'estimation de Bayes tend également vers zéro, et pour $\mu \to \infty$ la valeur asymptotique de cette variance est égale à σ_0^2.

Notons que la fonction $J[\hat{a} \mid x(t)]$ dans l'exemple envisagé est en vertu de (3.189), (3.187) et (3.190) égale à

$$J[\hat{a} \mid x(t)] = \int_{-\infty}^{\infty} (a - \hat{a})^2 W_1[a \mid x(t)]\, da = M_2\{a \mid x(t)\} =$$

$$= \frac{1}{\sqrt{2\pi}}\left(\frac{\sigma_0^2}{1 + \sigma_0^2 s_T}\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} (a - \hat{a})^2 \times$$

$$\times \exp\left[-\frac{1 + \sigma_0^2 s_T}{\sigma_0^2}(a - \hat{a})^2\right] da = \frac{\sigma_0^2}{1 + \sigma_0^2 s_T}, \tag{3.194}$$

c'est-à-dire qu'elle *ne dépend pas* de $x(t)$. Par conséquent, le risque moyen est égal à

$$R = m_1\{J[\hat{a} \mid x(t)]\} = \frac{\sigma_0^2}{1+\sigma_0^2 s_T} = \frac{\mu}{1+\mu}\cdot\frac{1}{s_T}. \qquad (3.195)$$

Pour $\sigma_0 \to 0$ le risque moyen tend également vers zéro car dans ce cas la répartition a priori du paramètre estimé tend vers une fonction delta.

Pour $\sigma_0 \to \infty$ on a

$$R \to R_{\max} = \frac{1}{s_T}, \qquad (3.196)$$

c'est-à-dire vers la valeur de la variance de l'estimation du maximum de vraisemblance.

Ainsi, la répartition a priori pour laquelle la variance est bien supérieure à celle de l'estimation du maximum de vraisemblance est la *moins favorable* et, par conséquent, l'estimation (3.166) du maximum de vraisemblance est également celle du *minimax*.

3.6.5. Propriétés asymptotiques de l'estimation de Bayes. Pour une répartition a priori arbitraire $w_1(a)$ de l'amplitude du'n signal quasi déterministe, l'estimation de Bayes de cette amplitude pour une fonction de pertes quadratique est en vertu de (3.186) égale à

$$\hat{a} = \int_{-\infty}^{\infty} aW_1[a \mid x(t)]\, da, \qquad (3.197)$$

où $x(t)$ est la réalisation observée sur l'intervalle $(-T, T)$ de la somme du signal $as(t)$ et d'un processus aléatoire normal de moyenne nulle et de fonction de corrélation donnée $B(t, y)$.

Portons dans (3.197) l'expression de la densité de probabilité a posteriori du paramètre estimé [cf. (3.187)], en utilisant l'expression pour la fonctionnelle du rapport de vraisemblance et les notations (3.188), (3.189); la formule (3.197) s'écrit alors comme suit:

$$\hat{a} = \frac{\int_{-\infty}^{\infty} a w_1(a) \exp\left(a x_T - \frac{a^2}{2} s_T\right) da}{\int_{-\infty}^{\infty} w_1(a) \exp\left(a x_T - \frac{a^2}{2} s_T\right) da}.$$

En complétant les exponentielles dans les expressions sous les intégrales de manière à avoir un carré entier, après des transformations algébriques simples on trouve

$$\hat{a} = \frac{\int_{-\infty}^{\infty} a w_1(a) \sqrt{\frac{s_T}{2\pi}} \exp\left[\frac{s_T}{2}\left(a - \frac{x_T}{s_T}\right)^2\right] da}{\int_{-\infty}^{\infty} w_1(a) \sqrt{\frac{s_T}{2\pi}} \exp\left[\frac{s_T}{2}\left(a - \frac{x_T}{s_T}\right)^2\right] da}. \qquad (3.198)$$

Si la variance de l'estimation du maximum de vraisemblance $s_T^{-1} \to 0$, on a

$$\sqrt{\frac{s_T}{2\pi}} \exp\left[\frac{s_T}{2}\left(a-\frac{x_T}{s_T}\right)^2\right] \to \delta\left(a-\frac{x_T}{s_T}\right).$$

La formule asymptotique pour l'estimation de Bayes de l'amplitude pour $s_T \to \infty$ s'obtient à partir de (3.198), supposant $w_1(a)$ continue et utilisant la propriété de filtrage de la fonction delta:

$$\hat{a} \sim \frac{x_T}{s_T} = \frac{\int_{-T}^{T} V(t)\, x(t)\, dt}{\int_{-T}^{T} V(t)\, s(t)\, dt}. \qquad (3.199)$$

Ainsi, quelle que soit la forme de la répartition a priori $w_1(a)$ de l'amplitude a, l'estimation de Bayes tend pour $s_T \to \infty$ vers l'estimation du maximum de vraisemblance (3.166).

Dans le cas où le processus additif est un bruit blanc on a

$$s_T = \frac{1}{N_0} \int_{-T}^{T} s^2(t)\, dt,$$

c'est-à-dire que la relation asymptotique mentionnée reste valable lorsque le rapport de l'énergie du processus déterministe à la densité spectrale du bruit blanc additif augmente indéfiniment.

Comme l'estimation non biaisée du maximum de vraisemblance (3.166) est répartie normalement et sa variance est s_T^{-1}, la densité de probabilité conditionnelle asymptotique (pour $s_T \to \infty$) de l'estimation de Bayes peut s'écrire comme suit:

$$W_1(\hat{a} \mid a) = \sqrt{\frac{s_T}{2\pi}} \exp\left[-\frac{s_T}{2}(\hat{a}-a)^2\right], \qquad (3.200)$$

par conséquent, la densité de probabilité non conditionnelle de l'estimation de Bayes $\hat{a}$ pour $s_T \to \infty$ tend asymptotiquement vers

$$W_1(\hat{a}) = \int_{-\infty}^{\infty} w_1(a)\, W_1(\hat{a} \mid a)\, da \sim w_1(\hat{a}), \qquad (3.200')$$

c'est-à-dire vers la densité de probabilité a priori du paramètre inconnu où la variable a est remplacée par l'estimation du maximum de vraisemblance de ce paramètre.

3.6.6. Généralisation des résultats. Revenons maintenant au cas général lorsque la composante déterministe $s(t;\ \vartheta_1, \ldots, \vartheta_m)$ contient un nombre quelconque fini de paramètres inconnus. Comme précédemment, la composante aléatoire est un processus aléatoire normal de moyenne nulle et de fonction de corrélation donnée.

Supposons que la fonction $s(t;\ \vartheta_1, \ldots, \vartheta_m)$ puisse s'écrire comme suit :

$$s(t;\ \vartheta_1, \ldots, \vartheta_m) = \sum_{j=1}^{m} \vartheta_j s_j(t), \tag{3.201}$$

où $s_j(t)$, $j = 1, \ldots, m$ sont des fonctions connues.

Calculons maintenant les estimations simultanées du maximum de vraisemblance des paramètres $\vartheta_1, \ldots, \vartheta_m$. En portant (3.201) dans le second membre de (3.158) on peut remplacer l'équation intégrale (3.158) par le système d'équations

$$\int_{-T}^{T} B(t, u) V_i(u)\, du = s_i(t), \qquad |t| \leqslant T,$$

$$i = 1, \ldots, m, \tag{3.202}$$

la fonction $V(t;\ \vartheta_1, \ldots, \vartheta_m)$ dont dépend le logarithme de la fonctionnelle du rapport de vraisemblance (3.157) est égale à

$$V(t;\ \vartheta_1, \ldots, \vartheta_m) = \sum_{j=1}^{m} \vartheta_j V_j(t). \tag{3.203}$$

Dérivons (3.203), il vient

$$\frac{\partial}{\partial \vartheta_i} V(t;\ \vartheta_1, \ldots, \vartheta_m) = V_i(t). \tag{3.204}$$

Portant (3.204) dans (3.161), obtenons le système d'équations du maximum de vraisemblance

$$\int_{-T}^{T} V_i(t) \left[x(t) - \sum_{j=1}^{m} \vartheta_j s_j(t) \right] dt = 0, \qquad i = 1, \ldots, m, \tag{3.205}$$

ou

$$\sum_{j=1}^{m} \vartheta_j \int_{-T}^{T} V_i(t) s_j(t)\, dt = \int_{-T}^{T} V_i(t) x(t)\, dt,$$

$$i = 1, \ldots, m. \tag{3.205'}$$

Introduisons les mêmes notations que dans (3.188) et (3.189), soit

$$s_{Tij} = \int_{-T}^{T} V_i(t) s_j(t)\, dt, \tag{3.206}$$

$$x_{Ti} = \int_{-T}^{T} V_i(t) x(t)\, dt. \tag{3.206'}$$

Ecrivons comme suit le système d'équations linéaires (3.205) par rapport à ϑ_j

$$\sum_{j=1}^{m} s_{Tij}\vartheta_j = x_{Ti}, \qquad i = 1, \ldots, m, \tag{3.207}$$

ou sous forme matricielle

$$\mathbf{s}_T \boldsymbol{\vartheta} = \mathbf{X}_T, \tag{3.207'}$$

où $\mathbf{s}_T$ est une matrice $m \times m$ dont les éléments sont s_{Tij}; $\boldsymbol{\vartheta}$ et $\mathbf{X}_T$ sont les vecteurs colonnes dont les éléments sont ϑ_j et x_{Ti} respectivement.

Supposant que pour tout j on ait

$$\int_{-T}^{T} s_j^2(t)\,dt < \infty$$

et que $B(t, u)$ soit une fonction définie positive (cf. § 3.3.1), on en déduit alors qu'il existe une matrice $\mathbf{s}_T^{-1}$ inverse de la matrice $\mathbf{s}_T$. La solution de l'équation (3.207') donne alors les estimations du maximum de vraisemblance des paramètres inconnus

$$\hat{\boldsymbol{\vartheta}} = \mathbf{s}_T^{-1}\mathbf{X}_T. \tag{3.208}$$

Si la composante aléatoire du processus observé est un bruit blanc de densité spectrale N_0, on a en vertu de (3.202)

$$N_0 V_i(t) = s_i(t), \qquad i = 1, \ldots, m,$$

et, par conséquent, les matrices $\mathbf{s}_T$ et $\mathbf{X}_T$ dans (3.207') deviennent (le facteur N_0 disparaît)

$$\mathbf{s}_T = \left\| \int_{-T}^{T} s_i(t)\, s_j(t)\,dt \right\|, \qquad i = 1, \ldots, m, \quad j = 1, \ldots, m, \tag{3.209}$$

$$\mathbf{X}_T = \left\| \int_{-T}^{T} s_i(t)\, x(t)\,dt \right\|, \qquad i = 1, \ldots, m, \tag{3.209'}$$

l'estimation vectorielle (3.208) est non biaisée. En effet, on obtient à partir de (3.208) *)

$$m_1\{\boldsymbol{\vartheta}\} = \mathbf{s}_T^{-1} m_1\{\mathbf{X}_T\}. \tag{3.210}$$

Mais

$$m_1\{x_{Ti}\} = \sum_{j=1}^{m} \vartheta_j \int_{-T}^{T} s_j(t)\, V_i(t)\,dt = \sum_{j=1}^{m} \vartheta_j s_{Tij},$$

par conséquent,

$$m_1\{\mathbf{X}_T\} = \mathbf{s}_T \boldsymbol{\vartheta}. \tag{3.210'}$$

*) Sur la moyenne des matrices voir § 2.7.1.

Portons (3.210') dans (3.210), il vient

$$m_1 \{\hat{\boldsymbol{\vartheta}}\} = \boldsymbol{\vartheta}, \tag{3.211}$$

ce qui montre que l'estimation (3.208) est non biaisée.

Nous allons trouver l'expression de la matrice de corrélation $\mathbf{M}$ des estimations (3.208) du maximum de vraisemblance [cf. (2.99)]

$$\mathbf{M} = m_1 \{(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})'\}, \tag{3.212}$$

où le signe « prime » indique qu'il s'agit d'un vecteur transposé (et dans la suite d'une matrice transposée). Après de simples réductions et substitution à $\hat{\boldsymbol{\vartheta}}$ de l'expression (3.208), nous obtenons compte tenu de la symétrie de la matrice $\mathbf{s}_T$:

$$\mathbf{M} = m_1 \{\hat{\boldsymbol{\vartheta}}\hat{\boldsymbol{\vartheta}}'\} - \boldsymbol{\vartheta}\boldsymbol{\vartheta}' = m_1 \{\mathbf{s}_T^{-1}\mathbf{X}_T\mathbf{X}_T' (\mathbf{s}_T^{-1})'\} - \boldsymbol{\vartheta}\boldsymbol{\vartheta}' = \\ = \mathbf{s}_T^{-1} m_1 \{\mathbf{X}_T\mathbf{X}_T'\} \mathbf{s}_T^{-1} - \boldsymbol{\vartheta}\boldsymbol{\vartheta}'. \tag{3.213}$$

Mais

$$m_1\{x_{Ti}x_{Tj}\} = \int_{-T}^{T}\int_{-T}^{T} B(u, v) V_i(u) V_j(v)\, du\, dv + \sum_{k=1}^{m} \vartheta_k s_{Tik} \sum_{n=1}^{m} \vartheta_n s_{Tjn},$$

par conséquent, [cf. (3.202)]

$$m_1 \{\mathbf{X}_T\mathbf{X}_T'\} = \mathbf{s}_T + \mathbf{s}_T\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\mathbf{s}_T. \tag{3.214}$$

Portons (3.214) dans (3.213), il vient

$$\mathbf{M} = \mathbf{s}_T^{-1}, \tag{3.215}$$

c'est-à-dire que la matrice de corrélation des estimations du maximum de vraisemblance de la composante déterministe (3.201) coïncide avec la matrice inverse de $\mathbf{s}_T$ [cf. (3.206) et pour un bruit blanc (3.209)].

Nous allons également calculer les éléments de la matrice d'information de Fisher [cf. (2.98)]. Utilisons le premier membre de (3.205), il vient

$$I^{(i,j)} = m_1 \left\{\frac{\partial \ln l}{\partial \vartheta_i} \frac{\partial \ln l}{\partial \vartheta_j}\right\} = m_1 \left\{ \int_{-T}^{T} V_i(u) \times \right. \\ \left. \times \left[x(u) - \sum_{k=1}^{m} \vartheta_k s_k(u)\right] du \int_{-T}^{T} V_j(v) \left[x(v) - \sum_{n=1}^{m} \vartheta_n s_n(v)\right] dv \right\} = \\ = \int_{-T}^{T}\int_{-T}^{T} B(u, v) V_i(u) V_j(v)\, du\, dv = \int_{-T}^{T} V_i(u) s_j(u)\, du,$$

ce qui coïncide avec (3.206).

Ainsi, la matrice d'information coïncide avec la matrice $\mathbf{s}_T$, c'est-à-dire

$$\mathbf{I} = \mathbf{s}_T, \tag{3.216}$$

et, utilisant (3.215),

$$\mathbf{I}^{-1} = \mathbf{s}_T^{-1} = \mathbf{M}. \tag{3.217}$$

Par conséquent [cf. (2.99)], les estimations du maximum de vraisemblance (3.208) sont *simultanément efficaces*.

Les résultats analogues peuvent être obtenus également quand les données des observations se présentent non pas sous la forme d'une réalisation $x(t)$, mais d'un échantillon discret (cf. problème 3.6).

On peut également généraliser les résultats du § 2.6.4 au cas des estimations simultanées de Bayes des paramètres aléatoires $\vartheta_1, \ldots, \vartheta_m$ d'une composante quasi déterministe de la forme (3.201) [cf. problème 3.7 ainsi que [11]).

3.6.7. Estimation des paramètres de la fonction de corrélation. Supposons qu'il y ait lieu d'estimer les paramètres inconnus $\vartheta_1, \ldots, \vartheta_m$ de la fonction de corrélation $B(t, y; \vartheta_1, \ldots, \vartheta_m)$ d'un processus aléatoire normal de moyenne nulle. Pour exclure la singularité nous allons supposer qu'on ajoute au processus un bruit blanc normal de densité spectrale N_0. Ainsi, la fonction de corrélation du processus aléatoire étudié $\xi(t)$ se présente sous la forme de la somme

$$B_\xi(t, y) = N_0\delta(y - t) + B(t, y; \vartheta_1, \ldots, \vartheta_m). \tag{3.218}$$

Nous allons utiliser l'expression (3.123) du logarithme de la fonctionnelle du rapport de vraisemblance en y posant

$$B_1(t, u) = B_\xi(t, u) \quad \text{et} \quad B_0(t, u) = N_0\delta(t-u),$$

$$\ln l[x(t); \vartheta_1, \ldots, \vartheta_m] = \frac{1}{2}\sum_{k=1}^{\infty}[\lambda_k(\vartheta_1, \ldots, \vartheta_m) - 1] \times$$

$$\times \left[\int_{-T}^{T} x(t)\varphi_k(t; \vartheta_1, \ldots, \vartheta_m)\,dt\right]^2 -$$

$$-\frac{1}{2}\sum_{k=1}^{\infty}\ln\lambda_k(\vartheta_1, \ldots, \vartheta_m), \tag{3.219}$$

où conformément à (3.120) φ_k et λ_k sont les nombres caractéristiques et les fonctions propres de l'équation intégrale linéaire

$$\int_{-T}^{T}[B(t, u; \vartheta_1, \ldots, \vartheta_m) + N_0\delta(t-u) - \lambda N_0\delta(t-u)]\,\varphi(u)\,du = 0$$

ou

$$N_0(\lambda-1)\varphi(t)=\int_{-T}^{T} B(t, u; \vartheta_1, \ldots, \vartheta_m)\varphi(u)\,du, \qquad |t|\leqslant T, \tag{3.220}$$

les fonctions propres étant normées conformément à (3.120′)

$$N_0\lambda_k(\vartheta_1, \ldots, \vartheta_m)\int_{-T}^{T}\varphi_k(t; \vartheta_1, \ldots, \vartheta_m)\varphi_n(t; \vartheta_1, \ldots, \vartheta_m)\,dt=$$

$$=\begin{cases} 1, & k=n, \\ 0, & k\neq n. \end{cases} \tag{3.221}$$

En prenant le maximum de la fonctionnelle du rapport de vraisemblance pour critère de qualité des estimations $\hat{\vartheta}_1, \ldots, \hat{\vartheta}_m$ des paramètres inconnus, nous trouvons que ces estimations doivent satisfaire au système d'équations

$$\frac{\partial}{\partial\vartheta_i}\ln l[x(t); \vartheta_1, \ldots, \vartheta_m]=0, \qquad i=1, \ldots, m, \tag{3.222}$$

où il y a lieu de substituer l'expression (3.219) pour $\ln l\,[x(t); \vartheta_1, \ldots, \vartheta_m]$.

Considérons plus en détail le cas où la fonction de corrélation $B(t, y)$ est donnée à un facteur constant près, c'est-à-dire

$$B(t, y) = cR(t, y). \tag{3.223}$$

Pour un processus aléatoire normal stationnaire, le paramètre inconnu c peut être la variance, alors $R(t, y) = R(y-t)$ est le coefficient de corrélation donné.

Dans ce cas l'équation intégrale (3.220) s'écrit

$$\varphi(t)=\mu\int_{-T}^{T} R(t, u)\varphi(u)\,du, \qquad |t|\leqslant T, \tag{3.224}$$

où

$$\mu=\frac{c}{N_0(\lambda-1)}. \tag{3.225}$$

Cette dernière formule donne pour λ

$$\lambda=1+\frac{c}{N_0\mu}. \tag{3.226}$$

Ecrivons maintenant la fonctionnelle du logarithme du rapport de vraisemblance. En portant (3.226) dans (3.219) et compte tenu

de la normalisation $\varphi_k(t;\, c) = \frac{1}{\sqrt{N_0\lambda_k}}\, \varphi_k(t)$ on obtient

$$\ln [x(t) \mid c] = \frac{1}{2N_0} \sum_{k=1}^{\infty} \frac{\frac{c}{N_0\mu_k}}{1+\frac{c}{N_0\mu_k}} \left[\int_{-T}^{T} x(t)\, \varphi_k(t)\, dt\right]^2 - \\ -\frac{1}{2}\sum_{k=1}^{\infty} \ln\left(1+\frac{c}{N_0\mu_k}\right). \quad (3.227)$$

En dérivant (3.227) par rapport à c, on trouve l'équation du maximum de vraisemblance

$$\frac{\partial}{\partial c} \ln l\,[x(t \mid c)] = \\ = \frac{1}{2N_0}\sum_{k=1}^{\infty}\left[\frac{1}{N_0\mu_k\left(1+\frac{c}{N_0\mu_k}\right)} - \frac{c}{N_0^2\mu_k^2\left(1+\frac{c}{N_0\mu_k}\right)^2}\right] \times \\ \times \left[\int_{-T}^{T} x(t)\, \varphi_k(t)\, dt\right]^2 - \frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{N_0\mu_k\left(1+\frac{c}{N_0\mu_k}\right)} = 0$$

ou

$$\frac{1}{N_0}\sum_{k=1}^{\infty}\frac{1}{\left(1+\frac{c}{N_0\mu_k}\right)^2\mu_k}\left[\int_{-T}^{T} x(t)\, \varphi_k(t)\, dt\right]^2 = \sum_{k=1}^{\infty}\frac{1}{\left(1+\frac{c}{N_0\mu_k}\right)\mu_k}. \quad (3.228)$$

En résolvant (3.228) par rapport à c, on obtient l'estimation du maximum de vraisemblance. Dans cette équation μ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale (3.224). Il est évident que le calcul de l'estimation du maximum de vraisemblance sous forme explicite à partir de (3.228) est très ardu. On remarque cependant que cette estimation s'obtient par des transformations *non linéaires* de la réalisation observée $x(t)$.

Nous allons maintenant trouver l'infimum de la variance des estimations non biaisées du paramètre c de la fonction de corrélation. A cet effet calculons l'information de Fisher [cf. (3.149)]:

$$I(c) = m_1\left\{\left(\frac{\partial}{\partial c}\ln l\,[x(t) \mid c]\right)^2\right\} = \\ = \frac{1}{4N_0^2}\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} m_1\left\{\left[\frac{\eta_k^2}{N_0\mu_k\left(1+\frac{c}{N_0\mu_k}\right)^2} - \frac{1}{\mu_k\left(1+\frac{c}{N_0\mu_k}\right)}\right]\times\right.$$

$$\times\left[\frac{\eta_n^2}{N_0\mu_n\left(1+\frac{c}{N_0\mu_n}\right)^2}-\frac{1}{\mu_n\left(1+\frac{c}{N_0\mu_n}\right)}\right]\Bigg\}=$$

$$=\frac{1}{4N_0^2}\sum_{k=1}^{\infty}\sum_{n=1}^{\infty}\frac{1}{\mu_k\mu_n\left(1+\frac{c}{N_0\mu_k}\right)^2\left(1+\frac{c}{N_0\mu_n}\right)^2}\times$$

$$\times m_1\left\{\left(\eta_k^2-N_0-\frac{c}{\mu_k}\right)\left(\eta_n^2-N_0-\frac{c}{\mu_n}\right)\right\},$$

où

$$\eta_k=\int_{-T}^{T}x(t)\,\varphi_k(t)\,dt$$

est une variable aléatoire normale de moyenne nulle et de variance $N_0+\frac{c}{\mu_k}$, η_k et η_n étant indépendants pour $k\neq n$. C'est pourquoi pour $k\neq n$ on a

$$m_1\left\{\left(\eta_k^2-N_0-\frac{c}{\mu_k}\right)\left(\eta_n^2-N_0-\frac{c}{\mu_n}\right)\right\}=0,$$

et pour $k=n$

$$m_1\left\{\left(\eta_k^2-N_0-\frac{c}{\mu_k}\right)^2\right\}=m_1\{\eta_k^4\}-\left(N_0+\frac{c}{\mu_k}\right)^2=2\left(N_0+\frac{c}{\mu_k}\right)^2.$$

Par conséquent

$$I(c)=\frac{1}{2N_0^2}\sum_{k=1}^{\infty}\frac{1}{\mu_k^2}\frac{1}{\left(1+\frac{c}{N_0\mu_k}\right)^2}. \tag{3.229}$$

En vertu de l'inégalité de Rao-Cramer [cf. (3.150)] on a pour la variance d'une estimation non biaisée

$$M_2\{\hat{c}\}\geqslant\frac{2N_0^2}{\sum\limits_{k=1}^{\infty}\frac{1}{\mu_k^2\left(1+\frac{c}{N_0\mu_k}\right)^2}}. \tag{3.230}$$

Compte tenu de (3.225), la limite inférieure de la variance [le second membre de l'inégalité (3.230)] s'écrira alors comme suit:

$$M_{2\min}=\frac{2N_0^2}{\sum\limits_{k=1}^{\infty}(\mu_k\lambda_k)^{-2}}. \tag{3.231}$$

Introduisons la fonction $h(t, y)$ définie par (3.132) et solution de l'équation intégrale [cf. (3.134)]

$$\int_{-T}^{T}R(t, y)\,h(v, y)\,dy+\frac{N_0}{c}h(v, t)=R(t, v),\quad |t|\leqslant T. \tag{3.232}$$

A partir de (3.231) on obtient l'expression de la limite inférieure de la variance de l'estimation du paramètre c sous la forme

$$M_{2\min}=\frac{2c^2}{\int\limits_{-T}^{T}\int\limits_{-T}^{T} h^2(t,\ y)\,dt\,dy}, \tag{3.233}$$

car en vertu de (3.132) on a

$$\int\limits_{-T}^{T}\int\limits_{-T}^{T} h^2(t,\ y)\,dt\,dy=\sum_{k=1}^{\infty}\left(\frac{\lambda_k-1}{\lambda_k}\right)^2=\frac{c^2}{N_0^2}\sum_{k=1}^{\infty}(\mu_k\lambda_k)^{-2}.$$

Si $\frac{N_0}{cT}\gg 1$, en qualité de solution approchée de l'équation intégrale (3.232) on peut prendre

$$h(v,\ t)\approx\frac{c}{N_0}R(t,\ v). \tag{3.234}$$

On aura alors pour la limite inférieure de la variance de l'estimaiton du paramètre c l'expression asymptotique suivante

$$\frac{M_{2\min}}{c^2}\sim\frac{2\left(\frac{N_0}{cT}\right)^2}{\frac{1}{T^2}\int\limits_{-T}^{T}\int\limits_{-T}^{T} R^2(u,\ v)\,du\,dv}. \tag{3.235}$$

Pour un processus stationnaire au sens général on a

$$\int\limits_{-T}^{T}\int\limits_{-T}^{T} R^2(u,\ v)\,du\,dv=\int\limits_{-2T}^{2T}(2T-|\tau|)\,R^2(\tau)\,d\tau$$

et la relation (3.235) s'écrira alors sous la forme

$$\frac{M_{2\min}}{c^2}\sim\left(\frac{N_0}{cT}\right)^2 k_T, \tag{3.235'}$$

où

$$\frac{1}{k_T}=\frac{1}{T}\int\limits_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)R^2(\tau)\,d\tau. \tag{3.236}$$

Notons qu'en utilisant (3.226) on peut écrire comme suit la formule (3.231):

$$\frac{M_{2\min}}{c^2}=\frac{2}{\sum\limits_{k=1}^{\infty}\left(1+\frac{N_0}{c}\mu_k\right)^{-2}}. \tag{3.236'}$$

Pour $N_0\to 0$ le dénominateur de (3.236') augmente indéfiniment (μ_k ne dépend pas de N_0), donc $M_{2\min}\to 0$. Ceci montre qu'il

est possible d'estimer sans biais, avec une variance nulle, un paramètre de la fonction de corrélation en l'absence de bruit blanc additif, ce qui correspond au cas singulier envisagé plus haut [cf. (3.130)].

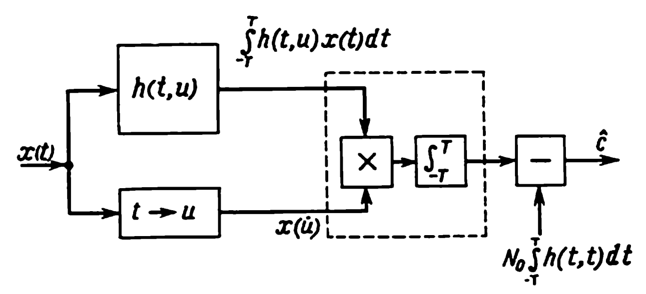


Fig. 3.1. Schéma d'un dispositif d'estimation du facteur d'amplitude d'une fonction de corrélation

Nous allons étudier maintenant une classe spéciale d'estimations non linéaires du paramètre inconnu c de la fonction de corrélation d'un processus aléatoire normal (fig. 3.1):

$$\hat{c}=\int_{-T}^{T}\int_{-T}^{T} h(t, u)\, x(t)\, x(u)\, dt\, du - N_0 \int_{-T}^{T} h(t, t)\, dt. \qquad (3.237)$$

Il est facile de voir que ces estimations sont *sans biais* si

$$\int_{-T}^{T}\int_{-T}^{T} h(t, u)\, R(t, u)\, dt\, du = 1, \qquad (3.238)$$

car compte tenu de (3.218) on a

$$m_1\{\hat{c}\}=\int_{-T}^{T}\int_{-T}^{T} h(t, u)\, B_\xi(t, u)\, dt\, du - N_0 \int_{-T}^{T} h(t, t)\, dt =$$
$$= c\int_{-T}^{T}\int_{-T}^{T} h(t, u)\, R(t, u)\, dt\, du = c.$$

La variance de l'estimation (3.237) est

$$M_2\{\hat{c}\}=m_1\{\hat{c}^2\}-c^2=$$
$$=\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T} h(t, u)\, h(v, y)\, m_1\{x(t)\, x(u)\, x(v)\, x(y)\}\times$$
$$\times\, dt\, du\, dv\, dy - 2N_0 \int_{-T}^{T} h(t, t)\, dt \int_{-T}^{T}\int_{-T}^{T} h(t, u)\, B_\xi(t, u)\, dt\, du +$$
$$+ N_0^2\left(\int_{-T}^{T} h(t, t)\, dt\right)^2 - c^2.$$

En utilisant l'expression bien connue du moment mixte du quatrième ordre d'un processus aléatoire normal (cf. problème 4.5, tome I), on obtient compte tenu de (3.218) et de la normalisation (3.238)

$$M_2\{\hat{c}\}=2N_0^2\int_{-T}^{T}\int_{-T}^{T}h(t, v)\,h(v, t)\,dt\,dv+$$

$$+4N_0c\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}h(t, u)\,h(t, y)\,R(u, y)\,du\,dy\,dt+$$

$$+c^2\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}h(t, u)\,h(v, y)\,[R(t, v)\,R(u, y)+$$
$$+R(t, y)\,R(u, v)]\,dt\,du\,dv\,dy. \qquad (3.239)$$

La recherche de la fonction de pondération $h(t, u)$ pour laquelle la variance $M_2\{\hat{c}\}$ est minimale pour la condition (3.238) est un problème compliqué du calcul des variations. Cependant, il est facile de résoudre ce problème dans les cas limites lorsque les rapports $\frac{N_0}{cT}$ sont petits et grands.

Si $\frac{N_0}{cT} \gg 1$, le problème formulé ci-dessus se réduit à minimiser la fonctionnelle [premier terme dans (3.239)]

$$\int_{-T}^{T}\int_{-T}^{T}h(t, v)\,h(v, t)\,dt\,dv$$

pour la condition (3.238). La solution est alors

$$h(t, u)=\frac{R(t, u)}{\int_{-T}^{T}\int_{-T}^{T}R^2(u, v)\,du\,dv}. \qquad (3.240)$$

Par conséquent, dans le cas envisagé on a

$$M_2\{\hat{c}\}\sim 2N_0^2\,\frac{\int_{-T}^{T}\int_{-T}^{T}R^2(t, v)\,dt\,dv}{\left(\int_{-T}^{T}\int_{-T}^{T}R^2(u, v)\,du\,dv\right)^2}=2N_0^2\left[\int_{-T}^{T}\int_{-T}^{T}R^2(u, v)\,du\,dv\right]^{-1}, \qquad (3.241)$$

ce qui en fait ne diffère pas de (3.235).

Ainsi, pour $\frac{N_0}{cT} \gg 1$ l'estimation non biaisée

$$\hat{c} = \frac{1}{\int\limits_{-T}^{T}\int\limits_{-T}^{T} R^2(u, v)\,du\,dv}\left[\int\limits_{-T}^{T}\int\limits_{-T}^{T} R(t, u)\,x(t)\,x(u)\,dt\,du - N_0\int\limits_{-T}^{T} R(t, t)\,dt\right] \qquad (3.242)$$

est une estimation *asymptotique efficace.*

Pour un processus normal stationnaire, (3.242) s'écrira

$$\hat{c} = \frac{1}{\int\limits_{-2T}^{2T}\left(1 - \frac{|\tau|}{2T}\right) R^2(\tau)\,d\tau}\left[\frac{1}{2T}\int\limits_{-T}^{T}\int\limits_{-T}^{T} R(t-u)\,x(t)\,x(u)\,dt\,du - N_0\right]. \qquad (3.243)$$

Dans l'autre cas limite lorsque $\frac{N_0}{cT} \to 0$ on a

$$M_2\{\hat{c}\} \to c^2 \int\limits_{-T}^{T}\int\limits_{-T}^{T}\int\limits_{-T}^{T}\int\limits_{-T}^{T} h(t, u)\,h(v, y)\times$$

$$\times\,[R(t, v)\,R(u, y) + R(t, y)\,R(u, v)]\,dt\,du\,dv\,dy > 0. \qquad (3.244)$$

Ainsi, dans la classe des estimations non biaisées considérées, la variance de l'estimation, même pour $\frac{N_0}{cT} \to 0$ (lorsque la singularité est possible), est limitée en bas par une grandeur différente de zéro.

3.6.8. Estimation de la fonction de corrélation. Dans les paragraphes précédents nous avons étudié les estimations des paramètres inconnus de la répartition d'un processus aléatoire. Dans certains cas, pour un type de répartitions donné, ce sont les fonctions déterministes caractérisant la répartition qui sont inconnues. Il s'agit avant tout de la fonction de corrélation et du spectre énergétique d'un processus aléatoire stationnaire (au sens général). Dans ce paragraphe et le suivant nous allons étudier les estimations de ces fonctions d'après une seule réalisation $x(t)$ du processus, observée sur l'intervalle $(-T, T)$.

Pour les processus aléatoires ergodiques la moyenne temporelle (cf. § 4.1.6, tome I)

$$B_T(\tau) = \frac{1}{2T}\int\limits_{-T}^{T} x(t)\,x(t+\tau)\,dt \qquad (3.245)$$

tend en probabilité pour $T \to \infty$ vers la fonction de corrélation $B(t)$ du processus aléatoire. Pour des hypothèses supplémentaires (par exemple, sur la normalité du processus), $B_T(\tau)$ converge en moyenne quadratique vers $B(\tau)$ (cf. problème 4.2, tome I).

Naturellement, on pourrait prendre pour l'estimation de la fonction de corrélation inconnue la fonction de corrélation temporelle échantillonnée (3.245). Cependant, cette fonction n'est pas commode car pour la calculer pour $|\tau| \leqslant T$ quelconque, il faut connaître des valeurs de la réalisation $x(t)$ à l'extérieur de l'intervalle d'observation $(-T, T)$. C'est pourquoi on prend pour l'estimation de la fonction de corrélation $B(\tau)$ la fonction

$$\hat{B}(\tau)=B_T^*(\tau)=\frac{1}{2T-|\tau|}\int\limits_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} x\left(t-\frac{\tau}{2}\right)x\left(t+\frac{\tau}{2}\right)dt, \quad |\tau|\leqslant T. \tag{3.246}$$

L'estimation (3.246) tout comme (3.245) est consistante et non biaisée. Cette dernière remarque est légitime car la moyenne sur un ensemble des réalisations est

$$m_1\{B_T^*(\tau)\}=\frac{1}{2T-|\tau|}\int\limits_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} m_1\left\{x\left(t-\frac{\tau}{2}\right)x\left(t+\frac{\tau}{2}\right)\right\}dt=B(\tau). \tag{3.247}$$

Pour calculer la variance de l'estimation (3.246) nous allons tout d'abord trouver le moment deux de la grandeur aléatoire $B_T^*(\tau)$:

$$m_2\{B_T^*(\tau)\}=\frac{1}{(2T-|\tau|)^2}\int\limits_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}}\int\limits_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} m_1\left\{x\left(u-\frac{\tau}{2}\right)\times\right.$$

$$\left.\times\, x\left(u+\frac{\tau}{2}\right)x\left(v-\frac{\tau}{2}\right)x\left(v+\frac{\tau}{2}\right)\right\}du\,dv. \tag{3.248}$$

Le processus étant supposé stationnaire, le moment quatre

$$m_1\left\{x\left(u-\frac{\tau}{2}\right)x\left(u+\frac{\tau}{2}\right)x\left(v-\frac{\tau}{2}\right)x\left(v+\frac{\tau}{2}\right)\right\}$$

dépend seulement des deux arguments τ et $v - u$. Désignant ce moment par $M(\tau, v-u)$ on peut écrire (3.248) comme suit *):

$$m_2\{B_T^*(\tau)\} = \frac{1}{(2T-|\tau|)^2} \int\limits_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} \int\limits_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} M(\tau, v-u)\, du\, dv =$$

$$= \frac{1}{(2T-|\tau|)^2} \int\limits_{-2T+|\tau|}^{2T-|\tau|} (2T-|\tau|-|z|)\, M(\tau, z)\, dz.$$

Comme

$$\frac{1}{(2T-|\tau|)^2} \int\limits_{-2T+|\tau|}^{2T-|\tau|} (2T-|\tau|-|z|)\, dz = 1,$$

et compte tenu de (3.247), on a

$$M_2\{B_T^*(\tau)\} = m_2\{B_T^*(\tau)\} - B^2(\tau) =$$

$$= \frac{1}{(2T-|\tau|)^2} \int\limits_{-2T+|\tau|}^{2T-|\tau|} (2T-|\tau|-|z|)\,[M(\tau, z) - B^2(\tau)]\, dz. \quad (3.249)$$

Pour un processus aléatoire *normal* stationnaire le moment quatre est égal à (cf. problème 4.5, tome I)

$$M(\tau, z) = B^2(\tau) + B^2(z) + B(\tau+z)\, B(\tau-z). \quad (3.250)$$

En portant (3.250) dans (3.249) on obtient l'expression de la variance de l'estimation de la fonction de corrélation d'un processus aléatoire normal

$$M_2\{B_T^*(\tau)\} = \frac{2}{2T-|\tau|} \int\limits_0^{2T-|\tau|} \left(1 - \frac{z}{2T-|\tau|}\right) \times$$

$$\times [B^2(z) + B(\tau+z)\, B(\tau-z)]\, dz. \quad (3.251)$$

On peut calculer la distribution des probabilités de l'estimation (3.246) par les méthodes du § 9.3 du tome I. Les résultats mentionnés dans ce paragraphe montrent en particulier que pour $T \to \infty$ cette estimation tend asymptotiquement vers la loi normale de moyenne $B(\tau)$ et de variance

$$M_2\{B_T^*(\tau)\} \sim \frac{1}{T} \int\limits_0^{\infty} [B^2(z) + B(\tau+z)\, B(\tau-z)]\, dz. \quad (3.252)$$

---

*) On passe de l'intégrale double à l'intégrale simple tout comme dans la transformation de (4.60) à (4.61) du tome I.

Pour $\tau = 0$ on trouve à partir de (3.246) l'estimation non biaisée de la puissance moyenne du processus aléatoire stationnaire

$$m_2^* = B_T^*(0) = \frac{1}{2T} \int_{-T}^{T} x^2(t)\,dt, \tag{3.253}$$

dont la variance, pour un processus normal, est selon (3.251) égale à

$$M_2\{m_2^*\} = \frac{2}{T} \int_0^{2T} \left(1 - \frac{z}{2T}\right) B^2(z)\,dz. \tag{3.254}$$

Pour $T \to \infty$ l'estimation (3.253) de la puissance moyenne tend asymptotiquement vers une loi normale de moyenne $B(0)$ et de variance

$$M_2\{m_2^*\} \sim \frac{2}{T} \int_0^{\infty} B^2(z)\,dz. \tag{3.255}$$

Notons que l'estimation (3.253) est un cas particulier de (3.237) pour $N_0 = 0$ et $h(t, u) = \frac{1}{2T}\,\delta(t - u)$. La formule (2.239) devient alors (2.254).

Les résultats ci-dessus peuvent facilement être généralisés à l'estimation de la fonction de corrélation mutuelle $B_{\xi\eta}(\tau)$ de deux processus $\xi(t)$ et $\eta(t)$ stationnaires et stationnairement liés. D'une manière analogue à (3.246) on a

$$B_{\xi\eta T}^*(\tau) = \frac{1}{2T - |\tau|} \int_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} x\left(t - \frac{\tau}{2}\right) y\left(t + \frac{\tau}{2}\right) dt, \quad |\tau| \leqslant T, \tag{3.256}$$

où $x(t)$ et $y(t)$ sont respectivement les réalisations des processus $\xi(t)$ et $\eta(t)$ observés sur l'intervalle $(-T, T)$.

L'estimation (3.256) est non biaisée, sa variance étant [comparer avec (3.249)]

$$M_2\{B_{\xi\eta T}^*(\tau)\} = \frac{1}{2T - |\tau|} \int_{-2T+|\tau|}^{2T-|\tau|} \left(1 - \frac{|z|}{2T - |\tau|}\right) [M_{\xi\eta}(\tau, z) - B_{\xi\eta}^2(\tau)]\,dz, \tag{3.257}$$

où

$$M_{\xi\eta}(\tau, z) = m_1\left\{x\left(u - \frac{\tau}{2}\right) y\left(u + \frac{\tau}{2}\right) x\left(v - \frac{\tau}{2}\right) y\left(v + \frac{\tau}{2}\right)\right\}. \tag{3.258}$$

3.6.9. Estimation du spectre énergétique. Estimons en première approximation le spectre énergétique d'un processus aléatoire con-

tinu *) stationnaire, d'après sa réalisation $x(t)$ sur l'intervalle $(-T, T)$, par la grandeur (cf. § 4.7.3, tome I) **)

$$\hat{F}(\omega) = G_T(\omega) = \frac{1}{T} |Z_T(i\omega)|^2, \tag{3.259}$$

où

$$Z_T(i\omega) = \int_{-T}^{T} x(t) e^{-i\omega t} dt. \tag{3.260}$$

Il est facile de voir que l'estimation (3.259) est *biaisée*. En utilisant la formule (4.61) du premier tome on a

$$m_1\{G_T(\omega)\} = F_T(\omega) = 4 \int_0^T \left(1 - \frac{\tau}{T}\right) B(\tau) \cos\omega\tau \, d\tau =$$

$$= F(\omega) - 4\left(\frac{1}{T}\int_0^T \tau B(\tau)\cos\omega\tau\, d\tau + \int_T^\infty B(\tau)\cos\omega\tau\, d\tau\right).$$

Par conséquent, le biais de l'estimation du spectre énergétique pour T donné est

$$F(\omega) - F_T(\omega) = \frac{4}{T}\int_0^T \tau B(\tau)\cos\omega\tau\, d\tau + 4\int_T^\infty B(\tau)\cos\omega\tau\, d\tau, \tag{3.261}$$

où $B(\tau)$ est la fonction de corrélation du processus.

Cependant, on conclut de (3.261) (comme d'ailleurs par définition du spectre énergétique) que l'estimation (3.259) est, pour $T \to \infty$, *asymptotiquement correcte*, en effet

$$\lim_{T\to\infty} F_T(\omega) = F(\omega),$$

pour

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \tau B(\tau)\, d\tau = 0.$$

Nous allons trouver la variance de l'estimation (3.259) du spectre énergétique d'un processus aléatoire *normal*. A cet effet nous allons tout d'abord calculer le moment deux

$$m_2\{G_T(\omega)\} = m_1\{G_T^2(\omega)\} = \frac{1}{T^2}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T} m_1\{x(t)\,x(y)\,x(u)\,x(v)\} \times$$
$$\times e^{-i\omega(t-y+u-v)}\, dt\, dy\, du\, dv.$$

*) On suppose par là même que la moyenne du processus est égale à zéro et qu'il ne contient pas de composantes quasi déterministes.

**) En statistique, l'estimation $G_T(\omega)$ est souvent appelée *périodogramme*.

Comme pour un processus normal on a (cf. problème 4.5, tome I)

$$m_1\{x(t)\,x(y)\,x(u)\,x(v)\} = B(y-t)\,B(v-u) + B(u-t)\,B(v-y) + B(v-t)\,B(u-y), \qquad (3.262)$$

alors (cf. (4.60), tome I)

$$m_2\{G_T(\omega)\} = \frac{1}{T^2}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}[B(y-t)\,B(v-u) + B(u-t)\,B(v-y) + B(v-t)\,B(u-y)]\,e^{-i\omega(t-y+u-v)}\times dt\,dy\,du\,dv = 2F_T^2(\omega) + \frac{1}{T^2}\int_{-T}^{T}\int_{-T}^{T}B(t-u)\times e^{-i\omega(t+u)}\,dt\,du\int_{-T}^{T}\int_{-T}^{T}B(v-y)\,e^{i\omega(y+v)}\,dy\,dv$$

ou

$$m_2\{G_T(\omega)\} = 2F_T^2(\omega) + \frac{1}{T^2}\left|\int_{-T}^{T}\int_{-T}^{T}B(u-t)\,e^{-i\omega(t+u)}\,dt\,du\right|^2. \qquad (3.263)$$

Utilisons (3.263) pour trouver la variance de l'estimation cherchée, on a

$$M_2\{G_T(\omega)\} = m_2\{G_T(\omega)\} - m_1^2\{G_T(\omega)\} = F_T^2(\omega) + \frac{1}{T^2}\left|\int_{-T}^{T}\int_{-T}^{T}B(u-t)\,e^{-i\omega(t+u)}\,dt\,du\right|^2. \qquad (3.264)$$

Comme pour $T\to\infty$ et $\omega\neq 0$ on a

$$\frac{1}{T^2}\left|\int_{-T}^{T}\int_{-T}^{T}B(u-t)\,e^{-i\omega(t+u)}\,dt\,du\right|^2 \sim F_T^2(\omega)\left(\frac{\sin 2\omega T}{2\omega T}\right)^2, \qquad (3.265)$$

le second terme dans (3.264) tend vers zéro pour $T\to\infty$ et $\omega\neq 0$, par conséquent

$$\lim_{T\to\infty} M_2\{G_T(\omega)\} = F^2(\omega). \qquad (3.266)$$

Pour $\omega = 0$ et $T\to\infty$ il découle directement de (3.264)

$$\lim_{T\to\infty} M_2\{G_T(0)\} = F^2(0).$$

Ainsi, pour tous les ω pour lesquels $F(\omega) > 0$, la variance de l'estimation du spectre énergétique *ne tend pas vers zéro* pour $T\to\infty$. Autrement dit, pour $T\to\infty$ l'estimation ne converge pas en moyenne quadratique vers le paramètre estimé.

Supposant toujours normal le processus étudié, on peut trouver la densité de probabilité de l'estimation (3.259) de son spectre éner-

gétique. Cette estimation est la somme des carrés des variables aléatoires normales

$$\frac{1}{\sqrt{T}}\int_{-T}^{T} x(t)\cos\omega t\,dt,\qquad \frac{1}{\sqrt{T}}\int_{-T}^{T} x(t)\sin\omega t\,dt,$$

dont les moyennes sont nulles, et les variances et la covariance sont respectivement

$$\sigma_{Tc}^2(\omega)=m_1\left\{\frac{1}{T}\left(\int_{-T}^{T} x(t)\cos\omega t\,dt\right)^2\right\}=$$

$$=\frac{1}{T}\int_{-T}^{T}\int_{-T}^{T} B(v-u)\cos\omega v\cos\omega u\,dv\,du, \qquad (3.267)$$

$$\sigma_{Ts}^2(\omega)=m_1\left\{\frac{1}{T}\left(\int_{-T}^{T} x(t)\sin\omega t\,dt\right)^2\right\}=$$

$$=\frac{1}{T}\int_{-T}^{T}\int_{-T}^{T} B(v-u)\sin\omega v\sin\omega u\,dv\,du, \qquad (3.268)$$

$$m_T(\omega)=m_1\left\{\frac{1}{T}\int_{-T}^{T} x(t)\cos\omega t\,dt\int_{-T}^{T} x(t)\sin\omega t\,dt\right\}=$$

$$=\frac{1}{T}\int_{-T}^{T}\int_{-T}^{T} B(v-u)\cos\omega v\sin\omega v\,dv\,du. \qquad (3.269)$$

Utilisons la formule (3.116) du tome I pour écrire l'expression de la fonction caractéristique de l'estimation du spectre énergétique, il vient :

$$\Theta_{G_T}(v)=\frac{1}{2\pi\sigma_{Tc}\sigma_{Ts}\sqrt{1-R_T^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-R_T^2)}\times\right.$$

$$\left.\times\left(\frac{x_1^2}{\sigma_{Tc}^2}-2R_T\frac{x_1x_2}{\sigma_{Tc}\sigma_{Ts}}+\frac{x_2^2}{\sigma_{Ts}^2}\right)+iv(x_1^2+x_2^2)\right\}dx_1\,dx_2=$$

$$=\frac{\sqrt{1-R_T^2}}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{\frac{1}{2}(y_1^2-2R_Ty_1y_2+y_2^2)+\right.$$

$$\left.+iv(1-R_T^2)(\sigma_{Tc}^2y_1^2+\sigma_{Ts}^2y_2^2)\right\}dy_1\,dy_2,$$

où l'on a introduit la notation

$$R_T=\frac{m_T}{\sigma_{Tc}\sigma_{Ts}}. \qquad (3.270)$$

Calculons l'intégrale d'après la méthode exposée dans l'annexe II. On a

$$\Theta_{G_T}(v)=\sqrt{1-R_T^2}\,[1-2iv\,(1-R_T^2)\,(\sigma_{Tc}^2+\sigma_{Ts}^2)-$$
$$-4v^2\,(1-R_T^2)^2\,\sigma_{Tc}^2\sigma_{Ts}^2-R_T^2]^{-\frac{1}{2}}$$

ou

$$\Theta_{G_T}(v)=[1-2iv\,(\sigma_{Tc}^2+\sigma_{Ts}^2)-4v^2\,(1-R_T^2)\,\sigma_{Tc}^2\sigma_{Ts}^2]^{-\frac{1}{2}}. \qquad (3.271)$$

Par transformation de Fourier inverse *) on peut trouver la densité de probabilité cherchée de l'estimation du spectre énergétique d'un processus aléatoire normal, soit:

$$W_{G_T}(z)=\frac{1}{2\sigma_{Tc}\sigma_{Ts}\sqrt{1-R_T^2}}\exp\left[-\frac{\sigma_{Tc}^2+\sigma_{Ts}^2}{4\sigma_{Tc}^2\sigma_{Ts}^2(1-R_T^2)}z\right]\times$$
$$\times I_0\left[\frac{\sqrt{(\sigma_{Tc}^2-\sigma_{Ts}^2)^2+4\sigma_{Tc}^2\sigma_{Ts}^2R_T^2}}{4\sigma_{Tc}^2\sigma_{Ts}^2(1-R_T^2)}z\right],\quad z\geqslant 0. \qquad (3.272)$$

Par définition du spectre énergétique, on a

$$\lim_{T\to\infty}\sigma_{Tc}^2(\omega)=\lim_{T\to\infty}\sigma_{Ts}^2(\omega)=\frac{1}{2}F(\omega) \qquad (3.273)$$

$$\lim_{T\to\infty}m_T(\omega)=0. \qquad (3.274)$$

En vertu de (3.272) pour $T\to\infty$ l'estimation du spectre énergétique tend à la limite vers une loi exponentielle

$$W_{G_T}(z)\sim\frac{1}{F(\omega)}e^{-\frac{z}{F(\omega)}},\quad z>0 \qquad (3.275)$$

de paramètre $F(\omega)$ [c'est-à-dire que la moyenne est égale à $F(\omega)$ et que la variance $F^2(\omega)$ correspond à (3.266)].

Bien que l'estimation $\hat{F}(\omega)=G_T(\omega)$ du spectre énergétique d'un processus aléatoire soit asymptotiquement correcte, elle a cet inconvénient que même en augmentant le temps d'observation de la réalisation du processus aléatoire, on n'arrive pas à atteindre la précision désirée car même pour $T\to\infty$, la variance de l'estimation reste finie. C'est pourquoi nous allons envisager également une classe d'estimations biaisées dont la variance croît avec l'augmentation du temps d'observation.

Estimons le spectre énergétique continu d'un processus aléatoire stationnaire par la grandeur aléatoire suivante, qui est une fonctionnelle de la réalisation observée du processus sur l'intervalle $(-T,\ T)$,

---

*) Voir, par exemple, [5], vol. 2, page 328.

soit :

$$\hat{F}(\omega) = g_T(\omega) = \frac{1}{T}\int_{-T}^{T}\left|\int_{-T}^{T} x(u)\,h(t-u)\,e^{-i\omega u}\,du\right|^2 dt, \quad (3.276)$$

où $h(\tau)$ est une fonction de pondération pouvant être interprétée comme la fonction de transition impulsionnelle d'un certain filtre linéaire. L'estimation (3.276) s'obtient en prenant la moyenne temporelle sur l'intervalle d'observation de la somme des carrés des processus à la sortie du filtre en question, lorsqu'à l'entrée on applique les processus $x(t)\cos\omega t$ et $x(t)\sin\omega t$.

En introduisant les transformées de Fourier $Z_T(i\omega)$ pour la réalisation $x(t)$ et $k(i\omega)$ pour la fonction de transition impulsionnelle $h(\tau)$ [cf. (5.5'), tome I] on peut écrire (3.276) comme suit :

$$g_T(\omega) = \frac{1}{(2\pi)^4}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{T}Z_T(i\omega_1)\times$$
$$\times Z_T(i\omega_2)\,k(i\omega_3)\,k(i\omega_4)\exp[iu(\omega_1-\omega_3-\omega)+$$
$$+iv(\omega_2-\omega_4+\omega)+it(\omega_3+\omega_4)]\,du\,dv\,dt\,d\omega_1\,d\omega_2\,d\omega_3\,d\omega_4. \quad (3.277)$$

Prenons la moyenne des deux membres de (3.277) sur l'ensemble des réalisations, passons à la limite pour $T\to\infty$ et utilisons la propriété de filtrage de la fonction delta

$$\delta(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega u}\,du,$$

(cf. annexe III); il vient

$$m_1\{g_T(\omega)\} \sim \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega+\nu)\,C^2(\nu)\,d\nu, \quad (3.278)$$

où

$$C^2(\omega) = |k(i\omega)|^2.$$

En examinant (3.278) on remarque que l'estimation (3.276) est biaisée. La caractéristique normée du filtre étant

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} C^2(\nu)\,d\nu = \int_{-\infty}^{\infty} h^2(\tau)\,d\tau = 1, \quad (3.279)$$

le biais asymptotique de l'estimation s'écrit comme suit

$$m_1\{g_T(\omega)\} - F(\omega) \sim \frac{1}{2\pi}\int_{-\infty}^{\infty}[F(\nu)-F(\omega)]\,C^2(\nu-\omega)\,d\nu. \quad (3.280)$$

En supposant de plus que le processus $x(t)$ soit normalement distribué, on peut trouver l'expression asymptotique (pour $T\to\infty$)

de la variance de l'estimation (3.276) [15], soit:

$$M_2\{g_T(\omega)\} \sim \frac{1}{T} F^2(\omega) \int_{-\infty}^{\infty} C^4(\omega)\, d\omega. \tag{3.281}$$

Pour un filtre de lissage du type RC-intégrateur (cf. § 5.2.3, tome I), compte tenu de (3.279), on a

$$C^2(\omega) = \frac{2/\alpha}{1+\left(\frac{\omega}{\alpha}\right)^2}, \quad \alpha = \frac{1}{RC},$$

et (3.281) donne

$$M_2\{g_T(\omega)\} \sim \frac{2\pi}{\alpha T} F^2(\omega), \tag{3.282}$$

ce qui signifie que pour obtenir une valeur suffisamment petite de la variance de l'estimation, le temps d'observation doit être bien supérieur à la constante de temps $\frac{1}{\alpha} = RC$ du filtre (autrement dit, le produit de la largeur de bande du filtre par le temps d'observation doit être bien supérieur à l'unité).

Problèmes

3.1. Soit $s(t)$ une fonction déterministe dont le spectre (transformée de Fourier) est continu et limité à la bande de fréquence (ω_1, ω_2). Montrer que dans ce cas on a la formule d'interpolation

$$s(t) = \sum_{n=-\infty}^{\infty} A\left(\frac{\pi n}{\Delta}\right) \frac{\sin\left[\Delta\left(t-\frac{\pi n}{\Delta}\right)\right]}{\Delta\left(t-\frac{\pi n}{\Delta}\right)} \cos\left[\omega_0\left(t-\frac{\pi n}{\Delta}\right) - \varphi\left(\frac{\pi n}{\Delta}\right)\right], \tag{1}$$

où

$$\Delta = \frac{\omega_2-\omega_1}{2}; \quad \omega_0 = \frac{\omega_2+\omega_1}{2}; \tag{2}$$

$$A^2(t) = s^2(t) + \sigma^2(t); \tag{3}$$

$$\varphi(t) = \operatorname{arc\,tg} \frac{\sigma(t)}{s(t)}; \tag{3'}$$

$\sigma(t)$ est la transformée de Hilbert de $s(t)$ (cf. annexe VI).

3.2. Montrer que la surface délimitée par une fonction déterministe $f(t)$ dont le spectre continu est identiquement nul pour $|\omega| > \Delta$ est égale à

$$\int_{-\infty}^{\infty} f(t)\, dt = \frac{\pi}{\Delta} \sum_{n=-\infty}^{\infty} f\left(\frac{\pi n}{\Delta}\right). \tag{4}$$

En utilisant (4) et la formule (4.67′) du premier tome montrer que le temps de corrélation d'un processus aléatoire stationnaire au sens général, dont le spectre énergétique est continu et limité à la bande de fréquence $(-\Delta, \Delta)$, est égal à

$$\tau_0 = \frac{\pi}{4\Delta} \sum_{n=-\infty}^{\infty} R\left(\frac{\pi n}{\Delta}\right), \tag{5}$$

où $R(\tau)$ est le coefficient de corrélation du processus.

3.3. Tenant compte du fait que le spectre d'amplitude de la fonction $\frac{\sin \Delta (t-t_0)}{\Delta (t-t_0)}$ s'annule pour $|\omega| \geqslant \Delta$ [cf. (3.11)] quel que soit t_0, démontrer l'égalité

$$\frac{\sin \Delta (t-t_0)}{\Delta (t-t_0)} = \sum_{n=-\infty}^{\infty} \frac{\sin \Delta \left(t_0 - \frac{\pi n}{\Delta}\right)}{\Delta \left(t_0 - \frac{\pi n}{\Delta}\right)} \frac{\sin \Delta \left(t - \frac{\pi n}{\Delta}\right)}{\Delta \left(t - \frac{\pi n}{\Delta}\right)}. \quad (6)$$

Conseil: utiliser le développement orthogonal de la fonction de corrélation (cf. § 3.3.2).

3.4. Démontrer que pour t_0 quelconque la formule (3.4) peut être généralisée comme suit:

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(t_0 + \frac{\pi n}{\Delta}\right) \frac{\sin \left[\Delta \left(t - t_0 - \frac{\pi n}{\Delta}\right)\right]}{\Delta \left(t - t_0 - \frac{\pi n}{\Delta}\right)}. \quad (7)$$

3.5. Montrer que l'estimation du maximum de vraisemblance de l'amplitude a d'un processus quasi déterministe $as(t)$, élaborée d'après la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$ de la somme de ce processus et d'un processus markovien normal stationnaire de variance σ^2 et de coefficient de corrélation $R(\tau) = \exp(-\mu |\tau|)$, est égale à

$$\hat{a} = \left\{\left[s(-T) - \frac{1}{\mu} s'(-T)\right] x(-T) + \left[s(T) - \frac{1}{\mu} s'(T)\right] x(T) + \right.$$
$$\left. + \mu \int_{-T}^{T} \left[s(t) - \frac{1}{\mu} s''(t)\right] x(t)\, dt\right\} \left[s^2(-T) + s^2(T) + \right.$$
$$\left. + \mu \int_{-T}^{T} \left\{s^2(t) + \frac{1}{\mu^2} [s'(t)]^2\right\} dt\right]^{-1}. \quad (8)$$

Montrer que l'estimation (8) est non biaisée et que sa variance [cf. (3.112)] est égale à

$$M_2\{\hat{a}\} = 2\sigma^2 \left[s^2(-T) + s^2(T) + \mu \int_{-T}^{T} \left\{s^2(t) + \frac{1}{\mu^2} [s'(t)]^2\right\} dt\right]^{-1}. \quad (9)$$

Etudier le cas particulier $s(t) \equiv 1$ et vérifier que l'on a

$$\hat{a} = \frac{x(-T) + x(T) + \mu \int_{-T}^{T} x(t)\, dt}{2(1 + \mu T)}, \quad (10)$$

$$M_2\{\hat{a}\} = \frac{\sigma^2}{1 + \mu T} \quad (11)$$

et pour $\mu T \gg 1$

$$\hat{a} \sim \frac{1}{2T} \int_{-T}^{T} x(t)\, dt, \quad (12)$$

$$M_2\{\hat{a}\} \sim \frac{\sigma^2}{\mu T}. \quad (13)$$

Expliquer pourquoi les formules (12) et (3.169) coïncident.

3.6. Etudier le problème § 3.6.5 en supposant que les résultats des observations sont donnés non pas par la somme $x(t) = s(t; \vartheta_1, \ldots, \vartheta_m) + \xi(t)$ mais par un échantillon discret $\mathbf{x} = (x_1, \ldots, x_N)$ où $x_i = x(t_i)$ et t_i sont les instants appartenant à l'intervalle d'observation $(-T, T)$. Montrer que les estimations du maximum de vraisemblance du paramètre ϑ_k peuvent s'écrire comme suit sous forme vectorielle [cf. (3.158)]:

$$\hat{\boldsymbol{\vartheta}} = \mathbf{S}^{-1}\mathbf{X}, \tag{14}$$

où

$$\mathbf{X} = \mathbf{s}'\mathbf{k}^{-1}\mathbf{x}, \tag{15}$$

$$\mathbf{S} = \mathbf{s}'\mathbf{k}^{-1}\mathbf{s}, \tag{15'}$$

$\mathbf{k}$ étant la matrice de corrélation normée de la composante aléatoire $\xi(t)$

$$\mathbf{k} = \frac{N\mathbf{K}}{\operatorname{tr}\mathbf{K}}, \tag{16}$$

où $\operatorname{tr}\mathbf{K}$ est la trace de la matrice de corrélation $\mathbf{K}$, et $\mathbf{s}$ la matrice $N \times m$ dont les colonnes sont linéairement indépendantes (au sens algébrique), la j-ième colonne de la matrice est le vecteur $\mathbf{s}_j$ de composantes $s_j(t_1), \ldots, s_j(t_N)$; $j = 1, \ldots, m$. Le signe « prime » désigne la matrice transposée. Montrer que l'estimation (14) est non biaisée et que sa matrice de corrélation est égale à

$$\mathbf{M} = \frac{\operatorname{tr}\mathbf{K}}{N}\mathbf{S}^{-1}. \tag{17}$$

3.7. Montrer que dans les conditions du problème 3.6 l'estimation simultanée de Bayes de l'ensemble des paramètres $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_m)$ pour une fonction de pertes quadratique peut s'écrire comme suit:

$$\hat{\boldsymbol{\vartheta}} = \frac{\int\limits_{\Omega} \boldsymbol{\vartheta} w_m(\boldsymbol{\vartheta})\, l_N(\mathbf{x} \mid \boldsymbol{\vartheta})\, d\boldsymbol{\vartheta}}{\int\limits_{\Omega} w_m(\boldsymbol{\vartheta})\, l_N(\mathbf{x} \mid \boldsymbol{\vartheta})\, d\boldsymbol{\vartheta}}, \tag{18}$$

où $w_m(\boldsymbol{\vartheta})$ est la densité de probabilité conjointe a priori des paramètres estimés et Ω le domaine de l'espace à m dimensions où cette fonction est donnée, la fonction $l_N(x \mid \boldsymbol{\vartheta})$ étant donnée par la relation

$$l_N(\mathbf{x} \mid \boldsymbol{\vartheta}) = \exp\left\{-\frac{N}{2\operatorname{tr}\mathbf{K}}(\mathbf{x} - \mathbf{s}\boldsymbol{\vartheta})'\,\mathbf{k}^{-1}(\mathbf{x} - \mathbf{s}\boldsymbol{\vartheta})\right\}. \tag{19}$$

3.8. Trouver l'infimum $M_{2\min}$ de la variance des estimations non biaisées du paramètre σ^2 d'un processus markovien normal stationnaire de fonction de corrélation (3.109), noyé dans un bruit blanc normal additif de densité spectrale N_0, les estimations étant obtenues d'après la réalisation du processus somme observée sur l'intervalle $(-T, T)$. Montrer que pour $\frac{N_0}{\sigma^2 T} \gg 1$ on a la formule asymptotique

$$\frac{M_{2\min}}{\sigma^4} \sim \frac{4\mu^2 t}{4\mu T + e^{-4\mu T} - 1}\left(\frac{N_0}{\sigma^2 T}\right)^2. \tag{20}$$

Montrer que lorsque le temps d'observation est très inférieur au temps de corrélation du processus markovien ($\mu T \ll 1$) on a en vertu de (20)

$$\frac{M_{2\min}}{\sigma^4} \sim \frac{1}{2}\left(\frac{N_0}{\sigma^2 T}\right)^2. \tag{20'}$$

Voir également le problème 7.6 dans le premier tome.

Chapitre 4

FILTRAGE DES PROCESSUS ALÉATOIRES

4.1. INTRODUCTION

Dans le chapitre précédent nous avons étudié les estimations des paramètres inconnus et des fonctions caractérisant la répartition des probabilités d'un processus aléatoire d'après une réalisation du processus observée sur un intervalle de temps fini. Il existe un autre problème ayant une portée pratique très grande, il s'agit de la recherche d'une transformation de la réalisation observée donnant la meilleure, en un certain sens, estimation de la valeur d'un processus aléatoire à un instant donné, à l'intérieur ou à l'extérieur de l'intervalle d'observation. Dans le cas général cette transformation est appelée *filtrage* du processus aléatoire.

Deux paragraphes de ce chapitre seront consacrés aux transformations linéaires. Laissant tout d'abord de côté la réalisation physique d'un filtre linéaire, on prend un intervalle de temps non limité, puis on restreint cet intervalle. Le critère de qualité de l'estimation de la valeur d'un processus aléatoire est dans certains cas la moyenne sur l'ensemble des réalisations du carré de l'erreur, c'est-à-dire le carré moyen de l'écart de l'estimation de la valeur estimée, et dans certains autres cas le rapport de la puissance du signal utile à la puissance du bruit. On cherche l'estimation *) des valeurs du processus aléatoire $\xi(t)$ d'après la réalisation observée $x(t)$ de la somme de ce processus et d'un autre processus $\eta(t)$, correlé avec $\xi(t)$. En se basant sur des résultats généraux on étudie quelques cas particuliers présentant un certain intérêt pour les applications radiotechniques. En plus de l'estimation du processus $\xi(t)$ il est également intéressant d'estimer, d'après la réalisation $x(t)$ observée sur un certain intervalle de temps fini (ou infini), des *transformations linéaires* de ce processus. On étudiera les modes de transformation linéaires suivants : déplacement de l'instant d'estimation sur l'axe des temps, dérivation et intégration simples et multiples et combinaisons de ces transformations.

Les questions de filtrage non linéaire d'après le critère de minimum d'erreur quadratique moyenne seront étudiées au § 4.4. On y considère une des approches possibles de ce problème où les systèmes non linéaires sont caractérisés par une somme infinie d'intégrales de Volterra. Par cette méthode on ne peut pas généralement

*) Parfois on dit que l'on *extrait* un processus (signal) $\xi(t)$ du mélange de ce processus et d'un autre processus (bruit) $\eta(t)$.

obtenir la solution exacte, c'est pourquoi on fait appel à des approximations qui dans certains cas peuvent, dans une large mesure, diminuer l'erreur quadratique moyenne correspondant à un système linéaire optimal.

Une autre approche du problème du filtrage non linéaire optimal consiste à limiter la classe des processus étudiés aux processus markoviens unidimensionnels ou aux composantes des processus markoviens multidimensionnels [13], [21]. Enfin, on utilise parfois des méthodes quasi linéaires en introduisant un paramètre de petite valeur (par exemple, le rapport signal/bruit défini d'une manière adéquate) [9], [10]. Ces deux méthodes ne sont pas étudiées dans le présent ouvrage et nous recommandons aux lecteurs intéressés de se référer aux ouvrages mentionnés ci-dessus.

4.2. FILTRAGE LINÉAIRE D'APRÈS LE CRITÈRE DE MINIMUM DU CARRÉ MOYEN DE L'ERREUR

4.2.1. Fonction de transition impulsionnelle d'un système linéaire optimal. Soit $x(t)$ une réalisation du processus somme $\xi(t)+\eta(t)$, définie pour toutes les valeurs réelles de t. Adoptons pour l'estimation de $\xi(t)$ la valeur filtrée de la réalisation (cf. § 5.2, tome I)

$$\hat{\xi}(t)=\int_{-\infty}^{\infty} h(t,\tau)\,x(\tau)\,d\tau, \tag{4.1}$$

où $h(t, \tau)$ est la fonction de transition impulsionnelle d'un filtre linéaire idéal.

Les intégrales des processus aléatoires sont ici et par la suite supposées convergentes en moyenne quadratique (cf. § 3.5, tome I).

Notons avant tout que l'estimation (4.1) est en général *biaisée* car

$$m_1\{\hat{\xi}(t)\}=\int_{-\infty}^{\infty} h(t,\tau)\,[a_\xi(\tau)+a_\eta(\tau)]\,d\tau\neq a_\xi(t),$$

où $a_\xi(t)$ et $a_\eta(t)$ sont les moyennes des processus $\xi(t)$ et $\eta(t)$.

La moyenne (sur l'ensemble des réalisations du processus aléatoire) du carré de l'erreur $\xi(t)-\hat{\xi}(t)$ pour cette estimation est égale à

$$\varepsilon^2(t)=m_1\{[\xi(t)-\hat{\xi}(t)]^2\}. \tag{4.2}$$

Calculons la fonction de transition impulsionnelle d'un filtre linéaire, qui présente la moindre erreur quadratique moyenne ε^2 parmi

tous les autres systèmes linéaires *).

En portant (4.1) dans (4.2) il vient après simples transformations

$$\varepsilon^2(t)=m_1\left\{\xi^2(t)-2\int_{-\infty}^{\infty} h(t,\tau)\,x(\tau)\,\xi(t)\,d\tau+\right.$$

$$\left.+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t,u)\,h(t,v)\,x(u)\,x(v)\,du\,dv\right\}=$$

$$=m_1\{\xi^2(t)\}-2\int_{-\infty}^{\infty} h(t,u)\,m_1\{x(\tau)\,\xi(t)\}\,d\tau+$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t,u)\,h(t,v)\,m_1\{x(u)\,x(v)\}\,du\,dv. \qquad (4.3)$$

Supposons que l'on connaisse les fonctions de corrélation $B_\xi(t_1, t_2)$ et $B_\eta(t_1, t_2)$ des processus $\xi(t)$ et $\eta(t)$ et leurs fonctions de corrélation mutuelles $B_{\xi\eta}(t_1, t_2)$, $B_{\eta\xi}(t_1, t_2)$. Comme

$$m_1\{x(u)\,x(v)\}=B_x(u,v)=B_\xi(u,v)+B_{\xi\eta}(u,v)+$$

$$+B_{\eta\xi}(u,v)+B_\eta(u,v), \qquad (4.4)$$

$$m_1\{x(\tau)\,\xi(t)\}=B_{x\xi}(\tau,t)=B_\xi(\tau,t)+B_{\xi\eta}(\tau,t), \qquad (4.5)$$

on a à partir de (4.3) à (4.5)

$$\varepsilon^2(t)=B_\xi(t,t)-2\int_{-\infty}^{\infty} h(t,\tau)\,B_{x\xi}(\tau,t)\,d\tau+$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t,u)\,h(t,v)\,B_x(u,v)\,du\,dv. \qquad (4.6)$$

On voit de (4.6) que l'erreur quadratique moyenne de l'estimation linéaire $\hat{\xi}(t)$ dépend seulement des fonctions de corrélation et de corrélation mutuelle des processus $\xi(t)$ et $\eta(t)$ et *ne dépend pas* de la structure plus fine de ces processus.

Nous allons montrer que si B_ξ, B_η, $B_{\xi\eta}$ et $B_{\eta\xi}$ sont données, le meilleur (au point de vue du critère adopté) filtrage linéaire du

*) Soulignons une fois de plus que l'on recherche l'optimum dans la classe des systèmes linéaires. Il est évident que dans certains cas, en utilisant des systèmes non linéaires, l'erreur peut être inférieure à celle que donnerait le système linéaire optimal. Si les processus $\xi(t)$ et $\eta(t)$ sont des processus aléatoires normaux, l'optimum absolu, comme nous allons le montrer, est toujours réalisé à l'aide d'un système linéaire. Le problème étudié ici du filtrage linéaire est une généralisation de la régression quadratique moyenne linéaire (cf. annexe XI) au cas des processus aléatoires.

processus $\xi(t)$ et de son mélange avec $\eta(t)$ sera réalisé par un système linéaire dont la fonction de transition impulsionnelle satisfait à l'équation intégrale *)

$$B_{x\xi}(\tau, t) = \int_{-\infty}^{\infty} h^*(t, y) B_x(\tau, y)\, dy. \tag{4.7}$$

Substituons (4.7) dans (4.6) et montrons que l'erreur $\varepsilon^2(t)$ est minimale si $h(t, \tau) \equiv h^*(t, \tau)$. En effet on a

$$\varepsilon^2(t) = B_\xi(t, t) - 2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t, \tau) h^*(t, y) B_x(\tau, y)\, d\tau\, dy +$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t, u) h(t, v) B_x(u, v)\, du\, dv =$$

$$= B_\xi(t, t) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(t, \tau) h^*(t, y) B_x(\tau, y)\, d\tau\, dy +$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B_x(u, v) [h(t, u) - h^*(t, u)] [h(t, v) - h^*(t, v)]\, du\, dv. \tag{4.8}$$

Comme seul le dernier terme de (4.8) contient la fonction inconnue $h(t, u)$ et qu'il n'est pas négatif (car la fonction de corrélation $B_x(u, v)$ est définie positive, cf. § 3.3.1), le minimum de ε^2 correspondra à un système linéaire dont la fonction de transfert annule cette erreur. Il est facile de voir que ceci a lieu si $h(t, u) \equiv h^*(t, u)$, ce qu'il fallait démontrer.

En vertu de (4.8), le minimum de l'erreur quadratique moyenne pour un système linéaire optimal est égal à

$$\varepsilon^2_{\min}(t) = B_\xi(t, t) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(t, u) h^*(t, v) B_x(u, v)\, du\, dv \tag{4.9}$$

ou, compte tenu de (4.7),

$$\varepsilon^2_{\min}(t) = B_\xi(t, t) - \int_{-\infty}^{\infty} h^*(t, u) B_{x\xi}(u, t)\, du. \tag{4.9'}$$

En utilisant la relation (5.14) du premier tome on peut écrire (4.9) comme suit:

$$\varepsilon^2_{\min}(t) = B_\xi(t, t) - B_{\hat{\xi}}(t, t), \tag{4.9''}$$

c'est-à-dire que le minimum du carré moyen de l'erreur est égal à la

*) Ici nous ne démontrons que la condition est suffisante, le fait qu'elle soit nécessaire est démontré, par exemple, dans [4].

différence des carrés moyens des valeurs du processus estimé et de l'estimation *).

La grandeur $\varepsilon^2_{\min}(t)$ peut également s'exprimer en fonction de l'intégrale de la différence des spectres énergétiques instantanés du processus $\xi(t)$ et de son estimation linéaire $\hat{\xi}(t)$. En vertu de (4.9″) on a

$$\varepsilon^2_{\min}(t) = \frac{1}{2\pi} \int_0^{\infty} [\Phi_{\xi}(t, \omega) - \Phi_{\hat{\xi}}(t, \omega)]\, d\omega,$$

où $\Phi(t, \omega)$ est le spectre énergétique instantané, l'indice de Φ désigne le processus auquel correspond le spectre.

Notons que pour les processus $\xi(t)$ et $\eta(t)$ non cohérents **) les équations

$$B_x(u, v) = B_{\xi}(u, v) + B_{\eta}(u, v), \tag{4.10}$$

$$B_{x\xi}(u, v) = B_{\xi}(u, v) \tag{4.10'}$$

et l'équation intégrale de base (4.7) deviennent

$$B_{\xi}(\tau, t) = \int_{-\infty}^{\infty} h^*(t, y)\,[B_{\xi}(\tau, y) + B_{\eta}(\tau, y)]\, dy, \tag{4.11}$$

et la formule (4.9) s'écrit alors

$$\varepsilon^2_{\min}(t) = B_{\xi}(t, t) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(t, u)\, h^*(t, v) \times$$

$$\times [B_{\xi}(u, v) + B_{\eta}(u, v)]\, du\, dv = B_{\xi}(t, t) - \int_{-\infty}^{\infty} h^*(t, u)\, B_{\xi}(u, t)\, du. \tag{4.11'}$$

Si les processus $\xi(t)$ et $\eta(t)$ sont stationnaires et stationnairement liés (du moins au sens général) et le filtre est un système linéaire de paramètres constants, l'équation intégrale (4.7) donnant la caractéristique de transition impulsionnelle s'écrit comme suit : ***)

$$B_{x\xi}(\tau) = \int_{-\infty}^{\infty} h^*(z)\, B_x(\tau - z)\, dz, \tag{4.12}$$

dans le cas où les processus $\xi(t)$ et $\eta(t)$ ne sont pas cohérents on a

$$B_{\xi}(\tau) = \int_{-\infty}^{\infty} h^*(z)\,[B_{\xi}(\tau - z) + B_{\eta}(\tau - z)]\, dz. \tag{4.12'}$$

---

*) En vertu de (4.9″) on a $B_{\hat{\xi}}(t, t) \leqslant B_{\xi}(t, t)$.

**) Ici on comprend la non-cohérence de deux processus aléatoires au sens défini page 158 du premier tome. Cette définition diffère également de la notion de « signal cohérent » ou de « signal non cohérent » souvent utilisée dans la technique des impulsions pour distinguer le cas où l'on connaît la phase du signal porteur H.F. de celui où cette phase est aléatoire.

***) L'équation intégrale (4.12) est souvent appelée équation de Wiener-Hopf.

La valeur du carré moyen de l'erreur ne dépend pas de l'instant où l'on procède à l'estimation de la valeur du processus $\xi(t)$, et son minimum, en vertu de (4.9), est égal à

$$\varepsilon_{\min}^2 = B_\xi(0) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(u)\,h^*(v)\,B_x(v-u)\,du\,dv =$$

$$= B_\xi(0) - \int_{-\infty}^{\infty} h^*(v)\,B_{x\xi}(v)\,dv \qquad (4.13)$$

ou

$$\varepsilon_{\min}^2 = B_\xi(0) - B_{\hat{\xi}}(0), \qquad (4.13')$$

c'est-à-dire que le minimum du carré moyen de l'erreur est égal à la différence des puissances moyennes du processus estimé et de l'estimation.

En exprimant les puissances moyennes en fonction des spectres énergétiques des processus [cf. (5.22), tome I] on peut écrire la grandeur $\varepsilon_{\min}^2$ comme suit:

$$\varepsilon_{\min}^2 = \frac{1}{2\pi}\int_0^{\infty} [F_\xi(\omega) - F_x(\omega)\,|\,k^*(i\omega)\,|^2]\,d\omega, \qquad (4.14)$$

$$F_x(\omega) = F_\xi(\omega) + F_\eta(\omega) + F_{\xi\eta}(\omega) + F_{\eta\xi}(\omega), \qquad (4.15)$$

où $k^*(i\omega)$ est la fonction de transfert d'un système linéaire optimal; $F_\xi(\omega)$, $F_\eta(\omega)$, $F_{\xi\eta}(\omega)$, $F_{\eta\xi}(\omega)$ sont les spectres énergétiques et les spectres énergétiques mutuels des processus $\xi(t)$ et $\eta(t)$.

Ainsi, le problème de la recherche du système linéaire optimal se réduit, dans le cas envisagé, à la solution de l'équation intégrale (4.12) qui, mise à part la possibilité de réalisation physique, s'obtient par transformation de Fourier des deux membres de (4.12). Comme le second membre de (4.12) est la convolution des fonctions dont les transformées de Fourier sont respectivement $k^*(i\omega)$ et $F_x(\omega)$, on a

$$F_{x\xi}(\omega) = F_\xi(\omega) + F_{\xi\eta}(\omega) = k^*(i\omega)\,F_x(\omega),$$

d'où compte tenu de (4.15) on trouve

$$k^*(i\omega) = \frac{F_{x\xi}(\omega)}{F_x(\omega)} = \frac{F_\xi(\omega) + F_{\xi\eta}(\omega)}{F_\xi(\omega) + F_\eta(\omega) + F_{\xi\eta}(\omega) + F_{\eta\xi}(\omega)}. \qquad (4.16)$$

En portant (4.16) dans (4.14) on obtient le minimum de l'erreur quadratique moyenne:

$$\varepsilon_{\min}^2 = \frac{1}{2\pi}\int_0^{\infty} \frac{F_\xi(\omega)\,F_x(\omega) - |\,F_{x\xi}(\omega)\,|^2}{F_x(\omega)}\,d\omega. \qquad (4.16')$$

Si les processus $\xi(t)$ et $\eta(t)$ ne sont pas cohérents, on a $F_{\xi\eta}(\omega) = F_{\eta\xi}(\omega) \equiv 0$ et en vertu de (4.16)

$$k^*(i\omega) = \frac{F_\xi(\omega)}{F_\xi(\omega) + F_\eta(\omega)}. \tag{4.17}$$

Comme le second membre de (4.17) est réel, il se trouve être la caractéristique de fréquence d'un filtre linéaire optimal (la caractéristique de phase est dans ce cas identiquement nulle).

En vertu de (4.16′) pour $F_{x\xi}(\omega) = F_\xi(\omega)$, $F_x(\omega) = F_\xi(\omega) + F_\eta(\omega)$ on obtient pour des processus non cohérents

$$\varepsilon^2_{\min} = \frac{1}{2\pi} \int_0^\infty \frac{F_\xi(\omega) F_\eta(\omega)}{F_\xi(\omega) + F_\eta(\omega)} d\omega. \tag{4.18}$$

La formule (4.18) montre que l'erreur dans un système linéaire optimal ne peut être nulle que si les spectres énergétiques des processus $\xi(t)$ et $\eta(t)$ ne se recouvrent pas, c'est-à-dire lorsque $F_\xi(\omega) F_\eta(\omega) \equiv 0$ pour toutes les valeurs de ω. Dans le cas contraire une erreur est inévitable. Pour qu'il n'y ait pas recouvrement il est évident qu'il faut que les spectres énergétiques $F_\xi(\omega)$ et $F_\eta(\omega)$ sur certains intervalles de l'axe des fréquences soient identiquement nuls *).

Ainsi, le filtrage d'un processus aléatoire stationnaire noyé dans un autre processus aléatoire stationnaire avec une erreur quadratique nulle (singularité) correspond au cas où les spectres énergétiques de ces processus ont des zones de fréquence où ils sont identiquement nuls.

Lors du filtrage du processus $\xi(t)$ noyé dans un bruit blanc de spectre $F_\eta(\omega) = N_0$, (4.18) donne

$$\varepsilon^2_{\min} = \frac{N_0}{2\pi} \int_0^\infty \frac{F_\xi(\omega)\, d\omega}{F_\xi(\omega) + N_0}, \tag{4.19}$$

c'est-à-dire que dans ce cas la présence d'un bruit blanc exclut toute singularité (cf. § 3.4.6).

4.2.2. Système linéaire optimal physiquement réalisable; temps d'observation fini. Un système linéaire dont la fonction de transition impulsionnelle $h^*(t, y)$ serait donnée par l'équation (4.7) est physiquement irréalisable, car $h^*(t, y)$ ne s'annule pas pour $t < y$. Les relations obtenues au paragraphe précédent peuvent être interprétées comme suit : le filtrage a lieu après l'observation de la réalisation $x(t)$ du processus somme $\xi(t) + \eta(t)$ sur tout l'intervalle de temps de $-\infty$ à ∞. Par conséquent, un tel système linéaire

*) Notons que l'un des spectres doit être une fonction à support borné. Leur rapport est alors égal soit à zéro, soit à l'infini [cf. (3.147)].

optimal permet d'estimer la valeur du processus à un instant donné avec un retard infini. La condition de réalisation physique signifie que lors du filtrage on utilise seulement la préhistoire de la réalisation $x(t)$ avant l'instant de l'estimation, autrement dit

$$h(t, \tau) \equiv 0, \quad \tau > t. \tag{4.20}$$

En portant (4.20) dans (4.1) on obtient l'estimation de $\xi(t)$ à l'aide d'un filtre linéaire physiquement réalisable

$$\hat{\xi}(t) = \int_{-\infty}^{t} h(t, \tau)\, x(\tau)\, d\tau. \tag{4.21}$$

Dans l'estimation (4.21) on utilise toutes les valeurs de la réalisation précédant l'instant t. Si on a une réalisation $x(\tau)$ de longueur finie T, c'est-à-dire si l'estimation à l'instant t s'effectue d'après les résultats des observations sur un intervalle fini $(t-T, t)$, on aura au lieu de (4.21)

$$\hat{\xi}(t) = \int_{t-T}^{t} h(t, \tau)\, x(\tau)\, d\tau = \int_{0}^{T} h(t, t-u)\, x(t-u)\, du. \tag{4.22}$$

En portant (4.22) dans (4.2) et après les mêmes transformations que dans (4.3), on obtient pour le carré moyen de l'erreur, compte tenu de la possibilité de réalisation physique du filtre et de la durée finie de la réalisation du processus, l'expression suivante:

$$\varepsilon^2(t) = B_\xi(t, t) - 2\int_0^T h(t, t-\tau)\, B_{x\xi}(t-\tau, t)\, d\tau + \\ + \int_0^T \int_0^T h(t, t-u)\, h(t, t-v)\, B_x(t-u, t-v)\, du\, dv. \tag{4.23}$$

En utilisant (4.23) il est facile (comme au § 4.2.1) de montrer que le meilleur, d'après le critère du minimum du carré moyen de l'erreur, filtrage de $\xi(t)$ du mélange de $\xi(t)$ et de $\eta(t)$ est assuré par un système linéaire dont la fonction de transition impulsionnelle satisfait à l'équation intégrale suivante:

$$B_{x\xi}(t-\tau, t) = \int_0^T h^*(t, t-y)\, B_x(t-\tau, t-y)\, dy, \quad 0 \leqslant \tau \leqslant T. \tag{4.24}$$

Le minimum du carré moyen de l'erreur correspondant au filtrage par un système linéaire optimale [cf. (4.9')] est égal à

$$\varepsilon^2_{\min}(t) = B_\xi(t, t) - \int_0^T h^*(t, t-u)\, B_{x\xi}(t-u, t)\, du. \tag{4.25}$$

Pour des processus stationnaires et stationnairement liés $\xi(t)$ et $\eta(t)$ et pour un filtre de paramètres constants, on a à partir de (4.24) et (4.25)

$$B_{x\xi}(\tau)=\int_0^T h^*(y)\,B_x(\tau-y)\,dy,\quad 0\leqslant\tau\leqslant T, \tag{4.26}$$

$$\varepsilon^2_{\min}=B_\xi(0)-\int_0^T h^*(u)\,B_{x\xi}(u)\,du. \tag{4.27}$$

Si dans ces formules on fait tendre T vers l'infini, on obtient le cas d'un temps d'observation infini précédant l'instant de l'estimation.

4.2.3. Estimation d'un processus après sa transformation linéaire. On peut maintenant étendre la méthode exposée ci-dessus à l'estimation des valeurs non pas du processus $\xi(t)$ mais de sa transformée $\zeta(t)$. Comme précédemment, l'estimation de $\zeta(t)$ s'effectue par filtrage linéaire de la réalisation $x(t)$ du mélange additif de $\xi(t)$ et de $\eta(t)$ observée sur un intervalle de temps fini $(t-T,\ t)$, c'est-à-dire

$$\hat{\zeta}(t)=\int_0^T h_1(t,\ t-u)\,x(t-u)\,du. \tag{4.28}$$

Le carré moyen de l'erreur pour une telle estimation est égal à

$$\varepsilon^2_\zeta(t)=m_1\left\{\left[\zeta(t)-\int_0^T h_1(t,\ t-u)\,x(t-u)\,du\right]^2\right\}=$$

$$=B_\zeta(t,\ t)-2\int_0^T h_1(t,\ t-\tau)\,B_{x\zeta}(t-\tau,\ t)\,d\tau+$$

$$+\int_0^T\int_0^T h_1(t,\ t-u)\,h_1(t,\ t-v)\,B_x(t-u,\ t-v)\,du\,dv, \tag{4.29}$$

où

$$B_{x\zeta}(t_1,\ t_2)=m_1\{x(t_1)\,\zeta(t_2)\}=B_{\xi\zeta}(t_1,\ t_2)+B_{\eta\zeta}(t_1\ t_2). \tag{4.29'}$$

La meilleure du point de vue du critère de minimum de ε^2_ζ estimation $\hat{\zeta}(t)$ s'obtient à l'aide d'un filtre linéaire, dont la fonction de transition impulsionnelle $h^*_1(t,\ \tau)$ satisfait à l'équation intégrale [cf. (4.24)]

$$B_{x\zeta}(t-\tau,\ t)=\int_0^T h^*_1(t,\ t-y)\,B_x(t-\tau,\ t-y)\,dy,\quad 0\leqslant\tau\leqslant T. \tag{4.30}$$

Le minimum du carré moyen de l'erreur est égal à

$$\varepsilon^2_{\zeta \min} = B_\zeta(t, t) - \int_0^T h_1^*(t, t-u)\, B_{x\zeta}(t-u, t)\, du. \tag{4.30'}$$

Pour des processus stationnaires les formules (4.30) et (4.30') s'écrivent comme suit:

$$B_{x\zeta}(\tau) = \int_0^T h_1^*(y)\, B_x(\tau - y)\, dy, \quad 0 \leqslant \tau \leqslant T, \tag{4.31}$$

$$\varepsilon^2_{\zeta \min} = B_\zeta(0) - \int_0^T h_1^*(u)\, B_{x\zeta}(u)\, du. \tag{4.31'}$$

Les différentes formes de transformations linéaires de $\xi(t)$ influent seulement sur la fonction de corrélation $B_\zeta(t_1, t_2)$ et la fonction de corrélation mutuelle $B_{x\zeta}(t_1, t_2)$. Nous allons donner ces expressions pour les trois types les plus importants de transformations linéaires.

Soit

$$\zeta(t) = \xi(t + t_0). \tag{4.32}$$

On a alors

$$B_\zeta(t, t) = B_\xi(t + t_0, t + t_0), \tag{4.33}$$

$$B_{x\zeta}(t_1, t_2) = B_\xi(t_1, t_2 + t_0) + B_{\eta\xi}(t_1, t_2 + t_0). \tag{4.33'}$$

Le cas $t_0 > 0$ *) ou $t_0 < -T$ correspond à l'*extrapolation* du processus aléatoire et le cas $-T \leqslant t_0 \leqslant 0$ à l'*interpolation*.

Si $\xi(t)$ est dérivable en moyenne quadratique et

$$\zeta(t) = \xi'(t), \tag{4.34}$$

on a (cf. § 4.3.3, tome I)

$$B_\zeta(t) = \left.\frac{\partial^2 B_\xi(t_1, t_2)}{\partial t_1\, \partial t_2}\right|_{t_1 = t_2 = t}, \tag{4.35}$$

$$B_{x\zeta}(t_1, t_2) = \frac{\partial B_\xi(t_1, t_2)}{\partial t_2} + \frac{\partial B_{\eta\xi}(t_1, t_2)}{\partial t_2}. \tag{4.35'}$$

Pour le processus

$$\zeta(t) = \int_{-\infty}^{\infty} g(t, \tau)\, \xi(\tau)\, d\tau, \tag{4.36}$$

où $g(t, \tau)$ est une fonction *donnée*, on obtient (cf. § 5.2.1, tome I)

$$B_\zeta(t, t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t, v_1)\, g(t, v_2)\, B_\xi(v_1, v_2)\, dv_1\, dv_2, \tag{4.37}$$

$$B_{x\zeta}(t_1, t_2) = \int_{-\infty}^{\infty} g(t_2, v)\, B_\xi(t_1, v)\, dv + \int_{-\infty}^{\infty} g(t_2, v)\, B_{\eta\xi}(t_1, v)\, dv. \tag{4.37'}$$

---

*) On appelle parfois le cas $t_0 > 0$ *pronostic*.

Dans ce cas l'estimation $\hat{\zeta}(t)$ [cf. (4.28)] faite d'après la réalisation de la somme $\xi(t) + \eta(t)$ observée sur un intervalle de temps fini, peut être interprétée comme une estimation optimale, du point de vue du critère de minimum du carré moyen de l'erreur, du processus à la sortie d'un système linéaire de fonction de transition impulsionnelle $g(t, \tau)$ donnée, lorsqu'on applique le processus $\xi(t)$ à son entrée.

Notons que pour $g(t, \tau) = \delta(\tau - t - t_0)$ à partir de (4.36)-(4.37') on obtient les formules (4.32)-(4.33') et pour $g(t, \tau) = \delta'(t - \tau)$ les formules (4.34)-(4.35').

On peut utiliser les formules (4.30) et (4.31) si l'on veut obtenir l'estimation linéaire optimale du processus $\zeta(t)$ qui est le résultat de la transformation de $\xi(t)$ dans un système linéaire de paramètres aléatoires. En utilisant la relation (5.112) du premier tome on peut écrire la fonction $B_\zeta(t, t)$ comme suit:

$$B_\zeta(t, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B_c(t, t, \omega_1, \omega_2) \times \times B_\xi(t-u, t-v)\, e^{i(\omega_1 u + \omega_2 v)}\, d\omega_1\, d\omega_2\, du\, dv, \qquad (4.38)$$

où

$$B_c(t_1, t_2, \omega_1, \omega_2) = m_1\{k(i\omega_1, t_1)\, k(i\omega_2, t_2)\} \qquad (4.38')$$

est la fonction de corrélation d'un système linéaire. On suppose que la fonction de transfert d'un système linéaire $k(i\omega, t)$, représentant le processus aléatoire, ne dépend pas de $\xi(t)$.

La fonction de corrélation mutuelle $B_{x\xi}(t_1, t_2)$ s'écrit dans ce cas comme suit:

$$B_{x\zeta}(t_1, t_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_c(i\omega, t_2) \times \times [B_\xi(t_1, t_2 - v) + B_{\xi\eta}(t_1, t_2 - v)]\, e^{i\omega v}\, d\omega\, dv, \qquad (4.39)$$

où

$$k_c(i\omega, t) = m_1\{k(i\omega, t)\} \qquad (4.39')$$

est la valeur moyenne de la fonction de transfert d'un système linéaire.

4.2.4. Le filtrage en tant que problème de régression. Le problème de filtrage linéaire étudié ci-dessus peut être interprété un peu différemment. Soient $\zeta(t)$ et $\xi(t)$ deux processus aléatoires *dépendants* et supposons qu'il y ait lieu d'estimer la valeur du processus $\zeta(t)$ par filtrage linéaire de la réalisation $x(t)$ du processus $\xi(t)$, en choisissant la caractéristique du filtre d'après le critère de minimum du carré moyen de l'erreur. La fonction de transition impulsionnelle de ce filtre peut être obtenue à partir de l'équation intégrale (4.30) [ou (4.31) pour des processus stationnaires].

Ainsi, le problème de filtrage est une généralisation au cas des processus aléatoires du problème de régression quadratique moyenne linéaire des variables aléatoires (cf. annexe XI). L'équation intégrale de Wiener-Hopf (4.31) est analogue à un système d'équations algébriques linéaires déterminant l'hyperplan de la régression quadratique moyenne dans l'espace des échantillons de dimension finie [cf. (7) dans l'annexe XI].

Sous une forme générale la fonctionnelle de la régression donne l'estimation du processus $\zeta(t)$ d'après la réalisation $x(t)$ observée du processus $\xi(t)$ comme une *moyenne conditionnelle*

$$\hat{\zeta}(t) = m_1 \{\zeta(t) \mid x(t)\}.$$

Cette estimation minimise le carré moyen de l'erreur $m_1 \{(\zeta - \hat{\zeta})^2\}$. La moyenne conditionnelle n'est pas toujours une fonctionnelle *linéaire* de la réalisation $x(t)$. Ce n'est que si les processus $\xi(t)$ et $\zeta(t)$ sont normaux (simultanément) que la fonctionnelle de la régression est toujours linéaire (cf. annexe XI ainsi que § 4.4.4).

4.2.5. Filtrage d'un signal quasi déterministe. Pour illustrer la théorie du filtrage optimal nous allons tout d'abord étudier le problème de filtrage d'un signal quasi déterministe $\xi(t) = as(t)$ dont la forme est déterminée par une fonction donnée $s(t)$ et l'amplitude a est aléatoire. Le signal est noyé dans un processus aléatoire $\eta(t)$ indépendant de a, de moyenne nulle et de fonction de corrélation $B(t_1, t_2)$. On demande de trouver la caractéristique d'un filtre linéaire donnant la meilleure (du point de vue du critère de minimum du carré moyen de l'erreur) estimation $\hat{\xi}(t) = \hat{a}s(t)$ d'après une réalisation du mélange additif $\xi(t) + \eta(t)$ observée sur l'intervalle $(t - T, t)$. A cet effet il faut résoudre l'équation intégrale (4.24) en y substituant au lieu de $B_{x\xi}$ et B_x les expressions suivantes [cf. (4.4) et (4.5)]:

$$B_{x\xi}(t - \tau, t) = m_1 \{a^2\} s(t - \tau) s(t), \tag{4.40}$$

$$B_x(t - \tau, t - y) = m_1 \{a^2\} s(t - \tau) s(t - y) + B(t - \tau, t - y). \tag{4.40'}$$

Introduisons la notation

$$m^2 = m_1 \{a^2\},$$

pour le moment deux de l'amplitude. Après la substitution mentionnée on obtient

$$m^2 s(t - \tau) s(t) = \int_0^T h^*(t, t - y) [m^2 s(t - \tau) s(t - y) + B(t - \tau, t - y)] \, dy, \quad 0 \leqslant \tau \leqslant T. \tag{4.41}$$

Cherchons la solution de l'équation intégrale (4.41) sous la forme suivante:

$$h^* (t, \tau) = \lambda s(t)\, v(\tau), \tag{4.42}$$

où $v(\tau)$ est la solution d'une autre équation intégrale *)

$$\int_{t-T}^{t} v(u)\, B(y, u)\, du = s(y), \quad t-T \leqslant y \leqslant t. \tag{4.43}$$

Portons (4.42) dans (4.41). Compte tenu de (4.43), il vient

$$m^2 = \lambda \left[m^2 \int_0^T v(t-y)\, s(t-y)\, dy + 1 \right],$$

ce qui permet de trouver la grandeur λ:

$$\lambda = m^2 \left(1 + m^2 \int_{t-T}^{t} v(z)\, s(z)\, dz\right)^{-1}. \tag{4.44}$$

Ainsi, le filtrage optimal du signal quasi déterministe $as(t)$ noyé dans le processus aléatoire $\eta(t)$ peut être réalisé par un système linéaire dont la fonction de transition impulsionnelle est

$$h^*(t, \tau) = \frac{m^2 s(t)\, v(\tau)}{1 + m^2 \int_{t-T}^{t} v(z)\, s(z)\, dz}, \tag{4.45}$$

où $v(\tau)$ est la solution de l'équation intégrale (4.43). L'estimation du signal à l'instant t d'après la réalisation $x(\tau)$ de son mélange additif avec $\eta(t)$ observée sur l'intervalle $(t-T, t)$ est

$$\hat{as}(t) = \int_0^T h^*(t, t-u)\, x(t-u)\, du = \frac{m^2 s(t) \int_{t-T}^{t} v(z)\, x(z)\, dz}{1 + m^2 \int_{t-T}^{t} v(z)\, s(z)\, dz}, \tag{4.46}$$

l'estimation de l'amplitude aléatoire du signal **) étant

$$\hat{a} = \frac{m^2 \int_{t-T}^{t} v(z)\, x(z)\, dz}{1 + m^2 \int_{t-T}^{t} v(z)\, s(z)\, dz}. \tag{4.46'}$$

*) Dans le chapitre précédent nous avons déjà rencontré plusieurs fois une équation du type (4.43).

**) Soulignons que l'estimation (4.46') est, suivant le critère du carré moyen de l'erreur, l'estimation optimale dans la classe des fonctionnelles linéaires de la réalisation observée $x(t)$. On ne fait ici *aucune* hypothèse concernant la distribution du bruit additif. Pour $m^2 \to \infty$ cette estimation coïncide avec l'estimation du maximum de vraisemblance de l'amplitude d'un signal déterministe noyé dans un bruit *normal* additif [cf. (3.166)]. Notons cependant que (4.46') coïncide avec l'estimation de Bayes (3.190) lorsque la fonction de pertes est quadratique et le bruit est normalement réparti, en posant de plus dans (3.190) $\sigma_0^2 = m^2$ et $a_0 = 0$.

L'estimation $\hat{a}$ est *biaisée*, car en vertu de (4.46′) on a

$$m_1\{\hat{a}\}=m_1\{a\}\frac{m^2\int\limits_{t-T}^{t}v(z)\,s(z)\,dz}{1+m^2\int\limits_{t-T}^{t}v(z)\,s(z)\,dz},\tag{4.47}$$

mais la relation (4.47) montre que pour $m^2\to\infty$ cette estimation est asymptotiquement correcte.

Trouvons le carré moyen de l'erreur en utilisant l'estimation (4.46). En vertu de (4.30′) et (4.40′) on a

$$\varepsilon_{\min}^2(t)=m^2s^2(t)-\lambda m^2\int\limits_0^T s^2(t)\,s(t-\tau)\,v(t-\tau)\,d\tau=$$

$$=m^2s^2(t)\left[1-\lambda\int\limits_{t-T}^{t}v(z)\,s(z)\,dz\right]$$

ou, compte tenu de (4.44),

$$\varepsilon_{\min}^2(t)=m_1\{(\hat{a}-a)^2s^2(t)\}=\frac{m^2s^2(t)}{1+m^2\int\limits_{t-T}^{t}v(z)\,s(z)\,dz}.\tag{4.48}$$

Il découle de (4.48)

$$m_1\{(\hat{a}-a)^2\}=\frac{m^2}{1+m^2\int\limits_{t-T}^{t}v(z)\,s(z)\,dz},\tag{4.49}$$

de plus, pour $m^2\to\infty$ le carré moyen de l'écart entre l'estimation $\hat{a}$ et l'amplitude estimée a du signal tend vers sa valeur minimale égale à $\left(\int\limits_{t-T}^{t}v(z)\,s(z)\,dz\right)^{-1}$. Le rapport

$$\frac{m_1\{(\hat{a}-a)^2\}}{m_1\{(\hat{a}-a)^2\}_{m^2=\infty}}=\mathrm{e}=\frac{m^2\int\limits_{t-T}^{t}v(z)\,s(z)\,dz}{1+m^2\int\limits_{t-T}^{t}v(z)\,s(z)\,dz},\tag{4.50}$$

qui est inférieur ou égal à l'unité, montre de combien le carré moyen de l'erreur, calculé dans l'hypothèse que le moment deux de l'amplitude soit fini, est inférieur à cette grandeur dans le cas où le moment deux est infini. En comparant (4.50) et (4.47) on voit que

$$m_1\{\hat{a}\}=\mathrm{e}m_1\{a\}.\tag{4.50′}$$

Dans le cas particulier du filtrage d'un signal quasi déterministe noyé dans un bruit blanc de densité spectrale N_0, on a

$$B(t,u)=N_0\delta(t-u),\quad v(\tau)=\frac{1}{N_0}s(\tau),$$

et à partir de (4.45), (4.48) on obtient

$$h^*(t, \tau) = \frac{\frac{m^2}{N_0} s(t) s(\tau)}{1 + \frac{m^2}{N_0} \int\limits_{t-T}^{t} s^2(z)\, dz}, \tag{4.51}$$

$$\varepsilon^2_{\min}(t) = \frac{m^2 s^2(t)}{1 + \frac{m^2}{N_0} \int\limits_{t-T}^{t} s^2(z)\, dz}. \tag{4.52}$$

Les relations ci-dessus sont évidemment vraies dans le cas du filtrage d'un signal déterministe, si par m on entend l'amplitude du signal déterministe. Tous les résultats se rapportant à l'estimation de l'amplitude inconnue (non aléatoire) d'un signal déterministe, mentionnés au § 3.6.1 peuvent être obtenus à partir des formules correspondantes de ce paragraphe en faisant $m^2 \to \infty$.

Notons qu'en vertu de (4.46) la *variance* de l'estimation d'un signal quasi déterministe, faite d'après la réalisation d'un mélange du signal et du bruit, est égale à la variance de la composante de bruit à la sortie du filtre, c'est-à-dire

$$M_2\{\hat{a}s(t)\} = \int\limits_0^T \int\limits_0^T h^*(t, t-u)\, h^*(t, t-v) \times$$

$$\times B(t-u, t-v)\, du\, dv = \lambda^2 s^2(t) \int\limits_{t-T}^{t} v(z)\, s(z)\, dz =$$

$$= m^2 s^2(t) \frac{m^2 \int\limits_{t-T}^{t} v(z)\, s(z)\, dz}{\left(1 + m^2 \int\limits_{t-T}^{t} v(z)\, s(z)\, dz\right)^2}. \tag{4.53}$$

En comparant (4.53) et (4.48) on trouve

$$\frac{M_2\{\hat{a}s(t)\}}{\varepsilon^2_{\min}(t)} = \frac{m^2 \int\limits_{t-T}^{t} v(z)\, s(z)\, dz}{1 + m^2 \int\limits_{t-T}^{t} v(z)\, s(z)\, dz} < 1, \tag{4.54}$$

ce rapport tendant vers l'unité pour $m^2 \to \infty$.

4.2.6. Extrapolation pure. Notre second exemple sera consacré au problème de l'extrapolation des valeurs du processus aléatoire $\xi(t)$ d'après une réalisation *de ce même processus* observée sur un intervalle de temps fini. On peut obtenir la solution de ce problème

particulier de la solution générale mentionnée au § 4.2.3, à condition de poser $\zeta(t) = \xi(t + t_0)$ et $\eta(t) \equiv 0$. La valeur extrapolée $\hat{\xi}(t + t_0)$ est obtenue par filtrage linéaire de la réalisation $x(t)$ du processus $\xi(t)$

$$\hat{\xi}(t+t_0) = \int_0^T h(t, t-u)\, x(t-u)\, du, \tag{4.55}$$

quant à la caractéristique de transition impulsionnelle du filtre, le meilleur du point de vue du critère de minimum du carré moyen de l'erreur, elle peut être obtenue à partir de l'équation intégrale suivante [cf. (4.30)]:

$$B_{\xi}(t-\tau, t+t_0) = \int_0^T h^*(t, t-y)\, B_{\xi}(t-\tau, t-y)\, dy, \quad 0 \leqslant \tau \leqslant T. \tag{4.56}$$

La valeur (minimale) correspondante du carré moyen de l'erreur de l'extrapolation est [cf. (4.30′)]

$$\varepsilon^2(t, t_0) = B_{\xi}(t+t_0, t+t_0) - \int_0^T h^*(t, t-u)\, B_{\xi}(t-u, t+t_0)\, du. \tag{4.57}$$

La caractéristique de transition impulsionnelle $h^*(\tau)$ du système linéaire optimal de paramètres constants, permettant d'obtenir la valeur extrapolée

$$\hat{\xi}(t+t_0) = \int_0^T h^*(u)\, x(t-u)\, du \tag{4.58}$$

d'un processus stationnaire (au sens général), est donnée par la solution de l'équation intégrale suivante:

$$B_{\xi}(\tau+t_0) = \int_0^T h^*(y)\, B_{\xi}(\tau-y)\, dy, \quad 0 \leqslant \tau \leqslant T, \tag{4.59}$$

qui est un cas particulier de l'équation (4.56). Le carré moyen de l'erreur est alors

$$\varepsilon^2_{\min}(t_0) = B_{\xi}(0) - \int_0^T h^*(u)\, B_{\xi}(u+t_0)\, du. \tag{4.60*}$$

*) Notons que pour une extrapolation pure ne tenant pas compte des possibilités de réalisation physique du filtre, l'intégration dans (4.58) à (4.60) s'effectue sur toutes les valeurs réelles de la variable. L'équation intégrale a alors la solution triviale $h^*(y) = \delta(y + t_0)$; l'estimation coïncide avec la valeur de la réalisation, donc $\hat{\xi}(t + t_0) = x(t + t_0)$ et le carré moyen de l'erreur est égal à zéro.

4.3. FILTRES ADAPTÉS

4.3.1. Filtrage linéaire d'après le critère de maximum du rapport signal/bruit. Aux paragraphes précédents la qualité du filtrage linéaire était déterminée par la valeur de l'erreur. Souvent il est important que le filtrage assure un rapport signal/bruit maximal même au prix d'une forte déformation du signal.

Soit $x(t)$ la réalisation de la somme d'un signal *déterministe* $s(t)$ et d'un bruit aléatoire $\eta(t)$ stationnaire (au sens général) de moyenne nulle et de fonction de corrélation donnée $B_\eta(\tau)$. On observe la réalisation sur l'intervalle de temps T. L'estimation du signal est donnée par la valeur filtrée de la réalisation

$$\hat{s}(t)=\int_0^T h(\tau)\,x(t-\tau)\,d\tau, \tag{4.61}$$

où $h(\tau)$ est la fonction de transition impulsionnelle à déterminer du système linéaire (de paramètres constants dans le temps). Pour le critère de qualité de l'estimation on prend le maximum du rapport signal/bruit à la sortie du filtre, ce rapport pouvant être facilement calculé pour un filtre linéaire (cf. § 6.3.1, tome I). En effet, la formule (4.61) peut s'écrire comme suit:

$$\hat{s}(t)=s_1(t)+\nu(t), \tag{4.62}$$

où

$$s_1(t)=\int_0^T h(\tau)\,s(t-\tau)\,d\tau, \tag{4.62'}$$

$$\nu(t)=\int_0^T h(\tau)\,\eta(t-\tau)\,d\tau. \tag{4.62''}$$

Le rapport signal/bruit est défini comme le rapport du carré du signal à la sortie $s_1^2(t_0)$, à un certain instant t_0, à la variance du bruit à la sortie du filtre $\sigma_\nu^2 = M_2\{\nu(t)\}$, c'est-à-dire

$$\frac{\text{signal}}{\text{bruit}}=\frac{s_1^2(t_0)}{\sigma_\nu^2}. \tag{4.63}$$

On appelle filtre *adapté* un filtre linéaire, optimal du point de vue du critère de maximum du rapport signal/bruit.

4.3.2. Fonction de transition impulsionnelle et fonction de transfert d'un filtre adapté. Calculons la fonction de transition impulsionnelle d'un système linéaire pour lequel le rapport signal/bruit est maximal. Désignons cette grandeur par μ_{max}. Dans ce cas, pour tout système linéaire de fonction de transition impulsionnelle $h(\tau)$,

on a l'inégalité suivante [cf. (4.62), (4.63)]

$$\mu_{\max}\sigma_{\nu}^{2} - s_{1}^{2}(t_0) = \mu_{\max}\int_0^T\int_0^T h(u)\,h(v)\,B_{\eta}(v-u)\,du\,dv - \left(\int_0^T h(\tau)\,s(t_0-\tau)\,d\tau\right)^2 \geqslant 0, \qquad (4.64)$$

l'égalité correspondant à un filtre adapté. Il est facile de vérifier que le premier membre de (4.64) s'annule lorsque $h(\tau) = h^*(\tau;\ t_0)$ satisfait à l'équation intégrale suivante:

$$\int_0^T h^*(\tau;\ t_0)\,B_{\eta}(t-\tau)\,d\tau = \mathrm{k}(t_0)\,s(t_0-t), \qquad 0 \leqslant t \leqslant T, \qquad (4.65)$$

où la constante $\mathrm{k}(t_0)$ est égale à

$$\mathrm{k}(t_0) = \frac{\int_0^T h^*(\tau;\ t_0)\,s(t_0-\tau)\,d\tau}{\mu_{\max}} = \frac{s_1^*(t_0)}{\mu_{\max}} = \frac{\sigma_{\nu}^{*2}}{s_1^*(t_0)} \qquad (4.66)$$

(l'astérisque signifie que les valeurs du signal s_1 et de la variance σ_{ν}^{2} du bruit se rapportent au processus à la sortie du filtre adapté).

Notons que si $h^*(\tau;\ t_0)$ est la solution de l'équation (4.65), la fonction $ch^*(\tau;\ t_0)$ satisfait également à cette équation, c étant ici une constante et la grandeur $\mu_{\max}$ restant inchangée. Ainsi, il suffit de calculer la fonction de transition d'un filtre adapté à un facteur constant (échelle) près.

Dans le cas du filtrage d'un signal déterministe noyé dans un bruit blanc, la fonction de transition impulsionnelle d'un filtre adapté peut s'exprimer explicitement si dans (4.65) on pose $B_{\eta}(t-\tau) = N_0\delta(t-\tau)$. En utilisant la propriété de filtrage de la fonction delta on obtient

$$h^*(t;\ t_0) = \frac{\mathrm{k}(t_0)}{N_0}\,s(t_0-t), \qquad 0 \leqslant t \leqslant T. \qquad (4.67)$$

Par conséquent, dans le cas envisagé la fonction de transition impulsionnelle d'un filtre adapté est proportionnelle à l'image par rapport à l'axe vertical passant par le point $t = t_0$ du segment du signal $s(t)$ sur l'intervalle $(t_0 - T,\ t_0)$ avec translation ultérieure de l'origine des coordonnées au point $t = t_0$ (fig. 4.1). Dans le cas où $t_0 = T$ la fonction de transition impulsionnelle est l'image du signal par rapport à la verticale divisant en deux l'intervalle $(0,\ T)$.

La valeur maximale du rapport signal/bruit est égale à [cf. (4.64)]

$$\mu_{\max} = \frac{\left[\int_0^T s^2(t_0-\tau)\,d\tau\right]^2}{N_0\int_0^T s^2(t_0-\tau)\,d\tau} = \frac{1}{N_0}\int_0^T s^2(t_0-\tau)\,d\tau, \qquad (4.68)$$

c'est-à-dire au rapport de l'énergie du signal sur l'intervalle (t_0 — — T, t_0) à la densité spectrale N_0 du bruit.

La fonction de transfert k^* ($i\omega$; t_0) d'un filtre adapté, dont la fonction de transition impulsionnelle est donnée par (4.67), est égale à

$$k^*(i\omega;\ t_0)=\int_{-\infty}^{\infty} h^*(t;\ t_0)\,e^{-i\omega t}\,dt=\frac{k(t_0)}{N_0}\int_{0}^{T} s(t_0-t)\,e^{-i\omega t}\,dt=$$

$$=\frac{k(t_0)}{N_0}\,e^{-i\omega t_0}\int_{t_0-T}^{t_0} s(t)\,e^{i\omega t}\,dt$$

ou

$$k^*(i\omega;\ t_0)=\frac{k(t_0)}{N_0}\,\overline{A_s(\omega;\ t_0,\ T)}\,e^{-i\omega t_0}, \tag{4.69}$$

où A_s (ω; t_0, T) est le spectre du segment du signal s (t) sur l'intervalle (t_0 — T, t_0), le trait désignant la grandeur complexe conjuguée.

Ainsi, lorsque l'on extrait un signal de son mélange avec un bruit blanc, la fonction de transfert d'un filtre adapté est proportionnelle au spectre complexe conjugué du signal tronqué.

On peut obtenir l'expression générale de la fonction de transfert du filtre adapté pour un spectre énergétique quelconque F_η (ω) du

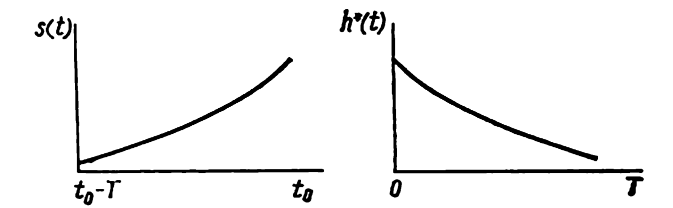


Fig. 4.1. Signal et fonction de transition impulsionnelle d'un filtre adapté

bruit si, renonçant à la condition de réalisation physique, on considère une réalisation x (t) pour toutes les valeurs réelles de t et on écrit le signal comme suit [comparer avec (4.61)]:

$$\hat{s}(t)=\int_{-\infty}^{\infty} h^*(\tau;\ t_0)\,x(t-\tau)\,d\tau.$$

Dans ce cas l'équation (4.65) devient (pour toutes les valeurs réelles de t)

$$\int_{-\infty}^{\infty} h^*(\tau;\ t_0)\,B_\eta(t-\tau)\,d\tau=k(t_0)\,s(t_0-t). \tag{4.70}$$

Après transformation de Fourier des deux membres de (4.70) on obtient l'expression

$$k^*(i\omega;\ t_0)\,F_\eta(\omega) = k(t_0)\,\overline{A_s(\omega)}\,e^{-i\omega t_0},$$

$A_s(\omega)$ étant le spectre du signal *non tronqué*, permettant de trouver la fonction de transfert du filtre adapté, soit:

$$k^*(i\omega;\ t_0) = \mathrm{k}(t_0)\,\frac{\overline{A_s(\omega)}}{F_\eta(\omega)}\,\mathrm{e}^{-i\omega t_0}. \tag{4.71}$$

Le rapport signal/bruit maximal est en vertu de (4.64) égal à (cf. également § 5.2.1, tome I)

$$\mu_{max} = \frac{\left[\int\limits_{-\infty}^{\infty} h^*(\tau,\ t_0)\,s(t_0-\tau)\,d\tau\right]^2}{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} h^*(u;\ t_0)\,h^*(v;\ t_0)\,B_\eta(v-u)\,du\,dv} = \frac{\left[\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} k^*(i\omega;\ t_0)\,A_s(\omega)\,e^{i\omega t}\,d\omega\right]^2}{\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} F_\eta(\omega)\,|k^*(i\omega;\ t_0)|^2\,d\omega}.$$

Substituant au lieu de $k^*(i\omega,\ t_0)$ son expression (4.71) on obtient finalement

$$\mu_{max} = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \frac{|A_s(\omega)|^2}{F_\eta(\omega)}\,d\omega. \tag{4.72}$$

Dans le cas d'un bruit blanc $F_\eta(\omega) = N_0$ et en vertu de (4.72) on a [comparer avec (4.68)]:

$$\mu_{max} = \frac{1}{2\pi N_0}\int\limits_{-\infty}^{\infty} |A_s(\omega)|^2\,d\omega. \tag{4.72'}$$

Notons que l'estimation du signal à la sortie d'un filtre adapté est en général *biaisée* car

$$m_1\{\hat{s}(t)\} = m_1\left\{\int\limits_0^T h^*(\tau;\ t_0)\,x(t-\tau)\,d\tau\right\} =$$

$$= \int\limits_0^T h^*(\tau;\ t_0)\,s(t-\tau)\,d\tau = s_1^*(t) \neq s(t). \tag{4.73}$$

La variance de cette estimation est égale à la variance de la composante de bruit $\nu(t)$ à la sortie du filtre adapté [cf. (4.53)]

$$M_2\{\hat{s}(t)\} = \sigma_\nu^{*2} \int_0^T \int_0^T h^*(u\,;\ t_0)\, h^*(v\,;\ t_0)\, B_\eta(v-u)\, du\, dv =$$

$$= \mathrm{k}(t_0) \int_0^T h^*(u\,;\ t_0)\, s(t_0-u)\, du = \mathrm{k}(t_0)\, s_1^*(t_0) = \frac{[s_1^*(t_0)]^2}{\mu_{\max}}. \qquad (4.73')$$

Pour un bruit blanc on a en vertu de (4.73), (4.73′) et (4.67)

$$m_1\{\hat{s}(t)\} = \frac{\mathrm{k}(t_0)}{N_0} \int_0^T s(t_0-\tau)\, s(t-\tau)\, d\tau,$$

$$M_2\{\hat{s}(t)\} = \frac{\mathrm{k}^2(t_0)}{N_0} \int_0^T s^2(t_0-\tau)\, d\tau.$$

4.3.3. Filtrage optimal d'une suite périodique d'impulsions noyées dans un bruit blanc. Nous allons illustrer par un exemple simple les résultats du paragraphe précédent. A cet effet nous allons calculer la caractéristique fréquentielle d'un filtre adapté destiné à extraire une suite périodique d'impulsions noyées dans un bruit blanc. Soit τ_0 la durée d'une impulsion et $u(t)$ une fonction différente de zéro sur l'intervalle $0 \leqslant t \leqslant \tau_0$ décrivant la forme de cette impulsion. Le spectre de l'impulsion est

$$A_u(\omega) = \int_0^{\tau_0} u(t)\, \mathrm{e}^{-i\omega t}\, dt.$$

Pour une suite d'impulsions de période T

$$s(t) = \sum_{n=-\infty}^{\infty} u(t-nT)$$

le spectre est

$$A_s(\omega) = \sum_{n=-\infty}^{\infty} A_u(\omega)\, \mathrm{e}^{-in\omega T} = \frac{A_u(\omega)}{1-\mathrm{e}^{-i\omega T}}.$$

La caractéristique fréquentielle normée du filtre adapté (c'est-à-dire le rapport du module du coefficient de transfert à $\frac{\mathrm{k}(t_0)}{N_0}$ aura en vertu de (4.71) la forme suivante:

$$C(\omega) = |A_s(\omega)| = \frac{|A_u(\omega)|}{2\left|\sin\frac{\omega T}{2}\right|}. \qquad (4.74)$$

Si la durée d'une impulsion tend vers zéro, la grandeur $|A_u(\omega)|$ reste constante dans une gamme de fréquence de plus en plus large et la caractéristique de fréquence du filtre adapté tend vers la caractéristique d'un *filtre en peigne* parfait (fig. 4.2)

$$C(\omega) \sim \frac{1}{2\left|\sin\frac{\omega T}{2}\right|}. \qquad (4.74')$$

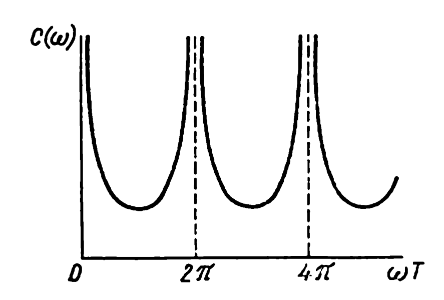


Fig. 4.2. Réponse en fréquence d'un filtre en peigne

Si le signal est un paquet d'impulsions de durée finie (contenant N impulsions) on a

$$A_s(\omega; N) = \sum_{n=0}^{N-1} A_u(\omega)\, e^{-in\omega T} = A_u(\omega)\frac{1-e^{-iN\omega T}}{1-e^{-i\omega T}}$$

et en vertu de (4.69) on a pour la caractéristique fréquentielle normée du filtre adapté

$$C(\omega) = |A_u(\omega)| \left|\frac{\sin\frac{N\omega T}{2}}{\sin\frac{\omega T}{2}}\right|. \qquad (4.75)$$

Notons qu'au contraire de la caractéristique fréquentielle (4.74) qui n'est pas limitée aux pulsations $\frac{2\pi k}{T}$, où k est un nombre entier quelconque et $\left|A_u\left(\frac{2\pi k}{T}\right)\right| > 0$, celle du filtre adapté, dans le cas d'une suite d'impulsions de durée finie, est limitée également aux pulsations multiples de la pulsation de cadence des impulsions, car en vertu de (4.75) on a

$$C\left(\frac{2\pi k}{T}\right) = N\left|A_u\left(\frac{2\pi k}{T}\right)\right|.$$

4.3.4. Filtre actif et filtre passif. Soit

$$\hat{s}(t) = \int_0^T h^*(\tau)\, x(t-\tau)\, d\tau \qquad (4.76)$$

l'estimation optimale, du point de vue du critère de maximum du rapport signal/bruit, du signal, où $h^*(\tau)$ est la solution de l'équation intégrale non homogène (4.76) et $x(t)$ une réalisation du mélange additif du signal et du bruit observée sur l'intervalle $(t - T, t)$. Le dispositif assurant une telle estimation peut être interprété de deux manières différentes. Selon l'une des méthodes, un filtre adapté séparant un signal d'un bruit aléatoire stationnaire, peut être inter-

prété comme un *système linéaire* de paramètres constants dont la fonction de transition impulsionnelle h^* (τ) est déterminée par la forme du signal et la fonction de corrélation du bruit [cf. (4.65)]. Dans le cas particulier du bruit blanc cette fonction de transition impulsionnelle s'obtient par réflexion du signal.

Cependant, le dispositif donnant à sa sortie l'estimation (4.76) peut être interprété différemment. Soit un oscillateur donnant la fonction h^* (τ) comme solution de l'équation intégrale (4.65). L'estimation $\hat{s}$ (t) est alors obtenue à l'aide d'un *corrélomètre* où la fonction h^* (τ) de l'oscillateur local se trouve multipliée par la réalisation reçue ayant traversée une ligne à retard, et le produit obtenu est alors intégré sur tout l'intervalle d'observation. Un tel dispositif de corrélation destiné à l'extraction du signal est parfois appelé *filtre actif* tandis qu'un système linéaire de fonction de transition impulsionnelle h^* (τ) est appelé *filtre passif* *).

4.4. FILTRAGE NON LINÉAIRE D'APRÈS LE CRITÈRE DE MINIMUM DE L'ERREUR QUADRATIQUE MOYENNE

4.4.1. Méthode générale permettant de caractériser des systèmes non linéaires. Dans les paragraphes précédents nous avons étudié les estimations d'un processus aléatoire, obtenues par filtrage linéaire de la réalisation observée, et les caractéristiques des filtres linéaires, optimaux vis-à-vis du critère de qualité adopté. Si l'on renonce à la condition de linéarité de l' algorithme de traitement de la réalisation observée, on peut, dans une classe plus large d'estimations, trouver des estimations qui pour le critère donné seront meilleures que les estimations linéaires.

Lorsque l'on utilise un système linéaire pour la recherche des estimations, la relation existant entre l'estimation et la réalisation observée est donnée par une équation intégrale assez simple du type (4.61). Si pour l'estimation on utilise un système inertiel non linéaire (avec des contre-réactions éventuelles), la relation simple entre l'estimation et la réalisation observée devient plus compliquée. En définissant un système non linéaire par une certaine équation intégro-différentielle non linéaire, on se heurte à des difficultés mathématiques importantes. Cependant, pour des systèmes linéaires, on peut tout de même essayer d'établir une relation explicite entre les processus d'entrée et de sortie.

La formule (9.5) du premier tome peut servir d'exemple de relation de ce genre, cette formule exprime le processus ζ (t) à la sortie d'un circuit type (amplificateur — détecteur quadratique — filtre) en fonction du processus d'entrée $x(t) = s(t) + \xi(t)$ sous la forme

*) Conformément à la terminologie de D. Middleton [10], des filtres actifs et passifs sont des filtres adaptés de première espèce, du premier et du second type respectivement.

d'une intégrale double, soit:

$$\zeta(t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(u,v)\,x(t-u)\,x(t-v)\,du\,dv, \qquad (4.77)$$

où

$$K(u,v)=\int_{-\infty}^{\infty}h_1(u-\tau)\,h_2(\tau)\,h_1(v-\tau)\,d\tau; \qquad (4.78)$$

$h_1(\tau)$, $h_2(\tau)$ étant respectivement les fonctions de transition impulsionnelles de l'amplificateur et du filtre.

Notons que pour le système non linéaire amplificateur-détecteur quadratique, c'est-à-dire pour $h_2(\tau) = \delta(\tau)$, le noyau $K(u, v)$ est égal au produit

$$K(u,v) = h_1(u)\,h_1(v). \qquad (4.78')$$

On peut généraliser (4.77) au cas où la caractéristique de l'élément non linéaire non intertiel du circuit type n'est pas parabolique, mais peut être approchée par un nombre fini de termes d'une série de puissances

$$z=\sum_{k=1}^{n}a_k y^k.$$

Au lieu de (4.77) on obtient alors

$$\zeta(t)=\sum_{k=1}^{n}a_k\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}K(u_1,\ldots,u_k)\,x(t-u_1)\ldots$$
$$\ldots x(t-u_k)\,du_1\ldots du_k, \qquad (4.79)$$

où

$$K(u_1,\ldots,u_k)=\int_{-\infty}^{\infty}h_1(u_1-\tau)\ldots h_1(u_k-\tau)\,h_2(\tau)\,d\tau. \qquad (4.80)$$

Pour $h_2(\tau) = \delta(\tau)$ le noyau multidimensionnel devient un produit de fonctions d'une variable, soit:

$$K(u_1,\ldots,u_k)=\prod_{i=1}^{k}h_1(u_i). \qquad (4.80')$$

Bien que souvent les schémas des systèmes non linéaires compliqués à contre-réaction puissent être remplacés par des circuits types étagés équivalents, nous indiquons la méthode générale permettant de caractériser un système non linéaire arbitraire. On part évidemment de l'hypothèse naturelle selon laquelle le processus à la sortie du système non linéaire est une fonctionnelle donnée sur l'ensemble des processus appliqués à son entrée. Selon un théorème remarquable

démontré en 1910 par le mathématicien français Fréchet *), pour toute fonctionnelle continue **) $y[x(t)]$ il existe une suite de fonctionnelles $y_n[x(t)]$ donnant pour $n \to \infty$ une approximation aussi bonne que l'on veut de $y[x(t)]$. La suite approximante s'écrit comme une somme d'intégrales (intégrales de Volterra), soit :

$$y_n[x(t)] = K_0 + \int_{-\infty}^{\infty} K_1(u_1)\, x(t-u_1)\, du_1 + \\ + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1, u_2)\, x(t-u_1)\, x(t-u_2)\, du_1\, du_2 + \ldots \\ \ldots + \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} K_n(u_1, \ldots, u_n)\, x(t-u_1)\ldots x(t-u_n)\, du_1\ldots du_n. \tag{4.81}$$

L'affirmation de Fréchet est un théorème d'existence, car dans la formule (4.81) les noyaux $K_m(u_1, \ldots, u_m)$, $m = 1, 2, \ldots$ restent non déterminés, et seule la possibilité de trouver dans chaque cas concret une suite de fonctionnelles approximantes se trouve garantie.

Le théorème de Fréchet est analogue au théorème bien connu de Weierstrass, se rapportant également aux théorèmes d'existence, sur l'approximation d'une fonction continue par une suite de polynômes. Généralement on réalise l'approximation des fonctions continues à l'aide d'un ensemble des polynômes orthogonaux. De cette façon on peut trouver les suites correspondantes de noyaux $K_m(u_1, \ldots, u_m)$ permettant d'approcher les fonctionnelles par la formule (4.81) (cf. § 4.4.5).

Notons que l'expression (4.79) du processus à la sortie d'un circuit type est un cas particulier de la relation (4.81) pour

$$K_m(u_1, \ldots, u_m) = K(u_1, \ldots, u_m),$$

où la fonction $K(u_1, \ldots, u_m)$ est donnée par (4.80).

Si $K_m \equiv 0$ pour tous les $m > 1$, on obtient une transformation purement linéaire de $x(t)$, $K_1(u_1)$ étant la fonction de transition impulsionnelle du filtre linéaire. En ajoutant des termes dans la série (4.81) pour $m > 1$, on introduit par là même une non-linéarité. L'ensemble des noyaux $K_0, K_1, \ldots, K_n$ caractérise un *filtre non linéaire de degré n*. Conformément à cette terminologie, les filtres linéaires sont des filtres du premier degré.

*) M. Fréchet, *Sur les fonctionnelles continues*. Ann. de l'Ecole Normale Supérieure, 3e Sér., 1910, p. 27.

**) Pour la notion de continuité dans l'espace des fonctionnelles voir, par exemple, le cours de G. Schilov, *Analyse mathématique*. Cours spécial (en russe). Physmathguiz, 1961, chap. 2, § 7.

Enfin, il faut noter que le théorème de Fréchet, en toute rigueur, concerne les fonctionnelles données sur un ensemble des fonctions continues déterministes. Sans entrer dans le détail, nous allons supposer que les conditions requises se trouvent remplies et que le théorème soit vrai même si le processus d'entrée du système non linéaire est aléatoire.

4.4.2. Filtres du second degré. Soit $x(t)$ une réalisation de la somme de deux processus aléatoires stationnaires et stationnairement liés $\xi(t)+\eta(t)$, déterminée pour toutes les valeurs réelles de t. Supposant que les moyennes des processus mentionnés soient nulles prenons pour l'estimation

$$\hat{\xi}(t)=\int_{-\infty}^{\infty} K_1(u_1)\,x(t-u_1)\,du_1+$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1,u_2)\,x(t-u_1)\,x(t-u_2)\,du_1\,du_2, \qquad (4.82)$$

où $K_1(u_1)=h^*(u_1)$ est la fonction de transition impulsionnelle du système linéaire optimal, donnée par l'équation (4.12).

A la différence des estimations étudiées au § 4.2, dans l'estimation (4.82) on voit apparaître un terme non linéaire. On voit s'ajouter ici à l'estimation optimale dans la classe des systèmes linéaires, un terme correcteur dû à la non-linéarité. Pour obtenir l'estimation (4.82) on a utilisé un filtre du second degré qui est un système non linéaire simple. Il s'agit de trouver la caractéristique $K_2(u_1,u_2)$ de la non-linéarité de telle sorte que le carré moyen de l'erreur

$$\varepsilon_2^2=m_1\{[\xi(t)-\hat{\xi}(t)]^2\} \qquad (4.83)$$

soit minimal.

Soit $\xi_1(t)$ l'erreur obtenue si pour estimer $\xi(t)$ on utilise seulement un système linéaire optimal, c'est-à-dire

$$\xi_1(t)=\xi(t)-\int_{-\infty}^{\infty} h^*(u_1)\,x(t-u_1)\,du_1. \qquad (4.84)$$

Notons que l'erreur $\xi_1(t)$ n'est pas corrélée avec $x(t)$ car, compte tenu de (4.12) on a

$$m_1\{x(t-v)\,\xi_1(t)\}=m_1\{x(t-v)\,\xi(t)\}-$$
$$-\int_{-\infty}^{\infty} h^*(u_1)\,m_1\{x(t-v)\,x(t-u_1)\}\,du_1=$$
$$=B_{x\xi}(v)-\int_{-\infty}^{\infty} h^*(u_1)\,B_x(u_1-v)\,du_1=0,$$

c'est-à-dire pour tout v

$$m_1 \{x(t-v)\,\xi_1(t)\} = 0. \tag{4.85}$$

Comme les processus sont stationnaires et stationnairement liés, en vertu de (4.82) à (4.84) on a

$$\varepsilon_2^2 = B_{\xi_1}(0) - 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1, u_2)\, m_{\xi_1 x}(u_1, u_2)\, du_1\, du_2 + \\ + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1, u_2)\times \\ \times K_2(u_3, u_4)\, m_x(u_1-u_2, u_1-u_3, u_2-u_4)\, du_1\, du_2\, du_3\, du_4, \tag{4.86}$$

où

$$m_{\xi_1 x}(u_1,u_2) = m_1\{\xi_1(t)\,x(t-u_1)\,x(t-u_2)\} = \\ = m_1\{\xi(t)\,x(t-u_1)\,x(t-u_2)\} - \\ - \int_{-\infty}^{\infty} h^*(u_3)\, m_1\{x(t-u_1)\,x(t-u_2)\,x(t-u_3)\}\, du_3 = \\ = m_{\xi x}(u_1,u_2) - \int_{-\infty}^{\infty} h^*(u_3)\, m_x(u_1-u_2, u_1-u_3)\, du_3\,; \tag{4.87}$$

$$m_x(u_1-u_2, u_1-u_3) = m_1\{x(t-u_1)\,x(t-u_2)\,x(t-u_3)\}\,; \tag{4.88}$$

$$m_x(u_1-u_2, u_1-u_3, u_2-u_4) = \\ = m_1\{x(t-u_1)\,x(t-u_2)\,x(t-u_3)\,x(t-u_4)\}, \tag{4.89}$$

et $B_{\xi_1}(0)$ coïncide avec le minimum de l'erreur quadratique moyenne donné par la formule (4.13), obtenue en utilisant des estimations linéaires.

Il découle de (4.86) que lorsque le dispositif d'estimation est un filtre non linéaire du second degré, l'erreur quadratique moyenne de l'estimation dépend non seulement des fonctions de corrélation et des fonctions de corrélation mutuelles des processus $\xi(t)$ et $\eta(t)$, mais également des moments mixtes du *troisième* et du *quatrième* ordre.

Pour trouver la condition imposée au noyau $K_2(u_1, u_2)$ du terme non linéaire de l'estimation et donnant une valeur minimale de ε_2^2, on procède de la même façon que dans le § 4.2.1. Montrons que si l'on se donne les moments mixtes des processus jusqu'au quatrième ordre inclus, le meilleur (vis-a-vis du critère adopté de minimum de l'erreur quadratique moyenne) filtrage non linéaire du second degré du processus $\xi(t)$ d'un mélange additif avec $\eta(t)$ aura lieu lorsque le noyau $K_2(u_1, u_2)$ du terme non linéaire correcteur satisfait à l'équation

intégrale suivante:

$$m_{\xi_1 x}(u_1, u_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2^*(u_3, u_4) \times$$
$$\times\, m_x(u_1-u_2, u_1-u_3, u_2-u_4)\, du_3\, du_4. \qquad (4.90)$$

Portons (4.90) dans (4.86) et montrons que l'erreur ε_2^2 est minimale si $K_2(u, v) = K_2^*(u, v)$. Nous avons après substitution

$$\varepsilon_2^2 = B_{\xi_1}(0) - 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1, u_2) \times$$
$$\times\, K_2^*(u_3, u_4)\, m_x(u_1-u_2, u_1-u_3, u_2-u_4)\, du_1\, du_2\, du_3\, du_4 +$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1, u_2)\, K_2(u_3, u_4) \times$$
$$\times\, m_x(u_1-u_2, u_1-u_3, u_3-u_4)\, du_1\, du_2\, du_3\, du_4 =$$
$$= B_{\xi_1}(0) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2^*(u_1, u_2)\, K_2^*(u_3, u_4) \times$$
$$\times\, m_x(u_1-u_2, u_1-u_3, u_2-u_4)\, du_1\, du_2\, du_3\, du_4 +$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [K_2(u_1, u_2) - K_2^*(u_1, u_2)]\,[K_2(u_3, u_4) -$$
$$- K_2^*(u_3, u_4)]\, m_x(u_1-u_2, u_1-u_3, u_2-u_4)\, du_1\, du_2\, du_3\, du_4. \qquad (4.91)$$

Nous avons utilisé ici l'égalité évidente [cf. (4.89)]

$$m_x(u_1 - u_2, u_1 - u_3, u_2 - u_4) = m_x(u_3 - u_4, u_3 - u_1, u_4 - u_2).$$

Comme seul le dernier terme dans (4.91) contient la fonction inconnue $K_2(u, v)$ qui n'est pas négative du fait que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [K_2(u_1, u_2) - K_2^*(u_1, u_2)]\,[K_2(u_3, u_4) -$$
$$- K_2^*(u_3, u_4)]\, m_x(u_1-u_2, u_1-u_3, u_2-u_4)\, du_1\, du_2\, du_3\, du_4 =$$
$$= m_1\left\{\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [K_2(u, v) - K_2^*(u, v)]\, x(t-u)\, x(t-v)\, du\, dv\right]^2\right\} \geqslant 0,$$

la valeur minimale de ε_2^2 correspond à un tel filtre non linéaire du second degré, pour lequel ce terme est identiquement nul. Ceci a lieu lorsque le noyau $K_2(u, v)$ du terme non linéaire correcteur est égal

à la fonction $K_2^*(u, v)$ qui est une solution de l'équation intégrale (4.90).

En posant dans (4.91) $K_2(u, v) \equiv K_2^*(u, v)$, cherchons la valeur minimale de l'erreur quadratique moyenne de l'estimation (4.82):

$$\varepsilon_{2\,\min}^2 = B_{\xi_1}(0) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2^*(u_1, u_2)\, K_2^*(u_3, u_4) \times$$

$$\times\, m_x(u_1-u_2,\ u_1-u_3,\ u_2-u_4)\, du_1\, du_2\, du_3\, du_4 \quad (4.92)$$

ou, compte tenu de (4.90),

$$\varepsilon_{2\,\min}^2 = \varepsilon_{1\,\min}^2 - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2^*(u_1, u_2)\, m_{\xi_1 x}(u_1, u_2)\, du_1\, du_2, \quad (4.93)$$

où

$$\varepsilon_{1\,\min}^2 = m_1\{\xi_1^2(t)\} = B_{\xi_1}(0) = B_{\xi}(0) - \int_{-\infty}^{\infty} h^*(u)\, B_{x\xi}(u)\, du$$

est l'erreur quadratique moyenne minimale de l'estimation linéaire [cf. (4.13)].

Ainsi, en utilisant un élément correcteur non linéaire dans un filtre du second degré on peut diminuer l'erreur quadratique moyenne, comparativement au cas de l'estimation par un système linéaire optimal seul, d'une grandeur égale à

$$\varepsilon_{2\,\min}^2 - \varepsilon_{1\,\min}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2^*(u_1, u_2)\, m_{\xi_1 x}(u_1, u_2)\, du_1\, du_2. \quad (4.94)$$

Etudions maintenant le problème d'un filtre optimal du second degré sous une forme plus générale en renonçant à l'hypothèse que le terme linéaire de l'estimation (4.82) soit donné. Nous allons trouver les *deux* fonctions $K_1(u)$ et $K_2(u, v)$ minimisant simultanément le carré moyen de l'erreur.

En portant (4.82) dans (4.83) nous trouvons l'expression de la fonctionnelle ε_2^2 qui dépend de deux noyaux inconnus du premier et du second ordre:

$$\varepsilon_2^2 = B_{\xi}(0) - 2\int_{-\infty}^{\infty} K_1(u_1)\, B_{\xi x}(u_1)\, du_1 -$$

$$-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1, u_2)\, m_{\xi x}(u_1, u_2)\, du_1\, du_2 +$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_1(u_1) K_1(u_2)\, B_x(u_2-u_1)\, du_1\, du_2 +$$

$$+2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_1(u_1)\,K_2(u_2,\ u_3)\,m_x(u_1-u_2,\ u_1-u_3)\,du_1\,du_2\,du_3+$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u_1,\ u_2)\,K_2(u_3,\ u_4)\times$$

$$\times m_x(u_1-u_2,\ u_1-u_3,\ u_2-u_4)\,du_1\,du_2\,du_3\,du_4. \qquad (4.95)$$

Tout comme précédemment, on peut montrer que le minimum du carré moyen de l'erreur correspond à un filtre non linéaire du second degré, dont les noyaux satisfont au *système de deux équations intégrales* suivant :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2^*(u_2,\ u_3)\,m_x(u_1-u_2,\ u_1-u_3)\,du_2\,du_3+$$

$$+\int_{-\infty}^{\infty} K_1^*(u_2)\,B_x(u_2-u_1)\,du_2=B_{\xi x}(u_1), \qquad (4.96)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2^*(u_3,\ u_4)\,m_x(u_1-u_2,\ u_1-u_3,\ u_2-u_4)\,du_3\,du_4+$$

$$+\int_{-\infty}^{\infty} K_1^*(u_3)\,m_x(u_1-u_2,\ u_1-u_3)\,du_3=m_{\xi x}(u_1,\ u_2). \qquad (4.96')$$

Notons que si les processus $\xi(t)$ et $\eta(t)$ sont indépendants et si leurs répartitions sont symétriques on a $m_x(u_1-u_2,\ u_1-u_3)\equiv 0$ et le système d'équations (4.96)-(4.96′) se décompose alors en deux équations dont la première devient (4.12) et la seconde (4.90) car dans ce cas $m_{\xi x}(u_1,\ u_2)=m_{\xi_1 x}(u_1,\ u_2)$. Par conséquent, lorsque la condition mentionnée se trouve réalisée, la solution obtenue ci-dessus pour le problème du filtre optimal du second degré est également vraie sans restrictions quelconques a priori imposées à la partie linéaire de l'estimation.

4.4.3. Filtres de degré quelconque. On peut améliorer l'estimation du processus $\xi(t)$ d'après la réalisation observée $x(t)$ de ce processus noyé dans un autre processus aléatoire $\eta(t)$ en utilisant des filtres non linéaires de degrés plus élevés. En vertu de (4.81) on peut écrire sous la forme suivante l'estimation donnée par un filtre de degré n:

$$\hat{\xi}(t)=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} K_i(u_1,\ \ldots,\ u_i)\times$$

$$\times x(t-u_1)\ldots x(t-u_i)\,du_1\ldots du_i. \qquad (4.97)$$

Le problème peut être posé ici de deux manières différentes. Tout d'abord on peut essayer de trouver une suite de noyaux

$K_1(u_1), \ldots, K_n(u_1, \ldots, u_n)$ minimisant le carré moyen de l'erreur, soit :

$$\varepsilon_n^2 = m_1\{[\xi(t) - \hat{\xi}(t)]^2\} = m_1\left\{\left[\xi(t) - \sum_{i=1}^{n} \int_{-\infty}^{\infty} \ldots \right.\right.$$
$$\left.\left. \ldots \int_{-\infty}^{\infty} K_i(u_1, \ldots, u_i)\, x(t-u_1) \ldots x(t-u_i)\, du_1 \ldots du_i\right]^2\right\}. \qquad (4.98)$$

La condition de minimum se réduit à un système de n équations intégrales par rapport aux noyaux inconnus, analogue au système (4.96)-(4.96′) pour $n = 2$.

Un problème moins général mais donnant un résultat relativement simple est analogue à celui qui a été mentionné au début du § 4.4.2. On suppose que $K_1(u)$ soit égal à la fonction de transition impulsionnelle d'un système linéaire optimal, le noyau $K_2(u_1, u_2)$ de l'élément correcteur non linéaire du second degré soit donné par l'équation (4.90) et le noyau $K_3(u_1, u_2, u_3)$ de l'élément correcteur du troisième degré se trouve de la condition de minimum de ε_3^2. En ajoutant une non-linéarité du quatrième degré on trouve $K_4(u_1, u_2, u_3, u_4)$ à partir de la condition de minimum de ε_4^2, etc. On a ainsi une équation de récurrence donnant $K_n^*(u_1, \ldots, u_n)$, si l'on connaît tous les noyaux optimaux jusqu'au degré $n - 1$ inclus.

Désignons par $\xi_{n-1}(t)$ l'erreur obtenue en utilisant pour l'estimation de $\xi(t)$ un filtre non linéaire de degré $n - 1$:

$$\xi_{n-1}(t) = \xi(t) - \sum_{i=1}^{n-1} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} K_i(u_1, \ldots, u_i)\, x(t-u_1) \ldots$$
$$\ldots x(t-u_i)\, du_1 \ldots du_i. \qquad (4.99)$$

On obtient alors à partir de (4.98)

$$\varepsilon_n^2 = m_1\left\{\left[\xi_{n-1}(t) - \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} K_n(u_1, \ldots, u_n) \times \right.\right.$$
$$\left.\left. \times\, x(t-u_1) \ldots x(t-u_n)\, du_1 \ldots du_n\right]^2\right\} = B_{\xi_{n-1}}(0) -$$
$$-2 \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} K_n(u_1, \ldots, u_n)\, m_{\xi_{n-1}x}(u_1, \ldots, u_n)\, du_1 \ldots du_n +$$
$$+ \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} K_n(u_1, \ldots, u_n)\, K_n(u_{n+1}, \ldots, u_{2n}) \times$$
$$\times\, m_x(u_1, \ldots, u_{2n})\, du_1 \ldots du_{2n}, \qquad (4.100)$$

où

$$m_{\xi_{n-1}x}(u_1, \ldots, u_n) = m_1\{\xi_{n-1}(t)\, x(t-u_1) \ldots x(t-u_n)\}; \qquad (4.101)$$
$$m_x(u_1, \ldots, u_{2n}) = m_1\{x(t-u_1) \ldots x(t-u_{2n})\}, \qquad (4.102)$$

et $B_{\xi_{n-1}}(0)$ coïncide avec l'erreur quadratique moyenne minimale pouvant être atteinte en utilisant les termes correcteurs non linéaires jusqu'au degré $n - 1$. Pour calculer le carré moyen de l'erreur il faut, en vertu de (4.100), connaître a priori les moments mixtes des processus initiaux d'ordre $2n$ inclus.

En procédant comme ci-dessus, il est facile de voir que le meilleur filtrage non linéaire de degré n, optimal vis-à-vis du critère de minimum du carré moyen de l'erreur, du processus $\xi(t)$ noyé dans un processus $\eta(t)$ est obtenu si le noyau $K_n(u_1, \ldots, u_n)$ satisfait à l'équation intégrale suivante:

$$m_{\xi_{n-1}x}(u_1, \ldots, u_n) = \\ = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} K_n^*(u_{n+1}, \ldots, u_{2n})\, m_x(u_1, \ldots, u_{2n})\, du_{n+1} \ldots du_{2n}. \tag{4.103}$$

Pour $n = 2$ l'équation (4.103) devient (4.90).

La valeur minimale de l'erreur quadratique moyenne de l'estimation donnée par un filtre non linéaire de degré n est égale

$$\varepsilon_{n\,\min}^2 = \varepsilon_{(n-1)\min}^2 - \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} K_n^*(u_1, \ldots, u_n) \times \\ \times m_{\xi_{n-1}x}(u_1, \ldots, u_n)\, du_1 \ldots du_n. \tag{4.104}$$

Pour $n = 2$ la formule (4.104) devient (4.93).

L'équation (4.103) est une relation de récurrence permettant en principe de trouver la caractéristique d'une non-linéarité correctrice de degré n optimale, connaissant les caractéristiques des éléments correcteurs optimaux de degré $n - 1$. La formule (4.104) donnant une réduction supplémentaire du carré moyen minimal de l'erreur grâce à l'introduction d'une non-linéarité de degré n, est également une formule de récurrence. Notons enfin que, tout comme dans la théorie du filtrage linéaire optimal, la méthode exposée ci-dessus peut être utilisée pour des cas plus généraux lorsque les processus $\xi(t)$ et $\eta(t)$ ne sont pas stationnaires, les réalisations de leur somme étant observées sur un intervalle de temps infini (ou semi-infini), l'estimation étant faite à partir d'une transformation linéaire du processus $\xi(t)$ et non de ce processus lui-même.

4.4.4. Filtrage d'un processus aléatoire normal. Nous n'avons jusqu'à présent fait aucune hypothèse spéciale quant à la loi de répartition des processus $\xi(t)$ et $\eta(t)$. Supposons maintenant que ces processus soient *normaux*. Il est alors facile de voir qu'en ajoutant des éléments non linéaires au filtre linéaire optimal on ne peut diminuer la valeur de l'erreur quadratique moyenne.

Comme nous l'avons déjà montré, il peut y avoir un filtrage non linéaire du second degré, optimal vis-à-vis du critère de minimum du

carré moyen de l'erreur, du processus $\xi(t)$ noyé dans un autre processus $\eta(t)$, si les caractéristiques $K_1(u)$ et $K_2(u, v)$ du filtre non linéaire vérifient de système d'équations intégrales (4.96), (4.96'). Pour des processus normaux on a $m_x(u_1 - u_2, u_1 - u_3) \equiv 0$, $m_{\xi x}(u_1, u_2) \equiv 0$ et, par conséquent, $K_2^*(u, v) \equiv 0$, donc, dans le sens indiqué, le meilleur filtrage est un filtrage linéaire.

Nous allons montrer maintenant qu'en ajoutant un élément non linéaire de degré quelconque, on n'arrive pas non plus à diminuer le carré moyen de l'erreur. Soit

$$\xi(t) - \hat{\xi}(t) = \xi_1(t) - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} K_n(u_1, \ldots, u_n) \times$$
$$\times x(t - u_1) \ldots x(t - u_n)\, du_1 \ldots du_n, \quad (4.105)$$

où $\xi_1(t)$ est l'erreur donnée par un système linéaire optimal [cf. (4.84)]. En répétant les raisonnements ayant abouti à (4.103), on voit que le minimum du carré moyen de l'erreur (4.105) peut être obtenu si le noyau $K_n(u_1, \ldots, u_n)$ satisfait à l'équation intégrale

$$m_{\xi_1 x}(u_1, \ldots, u_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} K_n^*(u_{n+1}, \ldots, u_{2n}) \times$$
$$\times m_x(u_1, \ldots, u_{2n})\, du_{n+1} \ldots du_{2n}, \quad n > 1. \quad (4.106)$$

Comme nous l'avons mentionné plus haut, l'erreur $\xi_1(t)$ n'est pas corrélée à $x(t)$. Pour des processus aléatoires normaux $\xi_1(t)$ et $x(t)$ ceci signifie leur indépendance. C'est pourquoi dans ce cas

$$m_{\xi_1 x}(u_1, \ldots, u_n) = m_1\{\xi_1(t)\, x(t - u_1) \ldots x(t - u_n)\} =$$
$$= m_1\{\xi_1(t)\}\, m_1\{x(t - u_1) \ldots x(t - u_n)\},$$

et comme

$$m_1\{\xi_1(t)\} = \int_{-\infty}^{\infty} h^*(u)\, m_1\{x(t - u)\}\, du = 0,$$

pour des processus normaux de moyenne nulle on a

$$m_{\xi_1 x}(u_1, \ldots, u_n) \equiv 0. \quad (4.107)$$

En vertu de (4.106) et de (4.107) on a

$$K_n^*(u_1, \ldots, u_n) \equiv 0, \ n > 1. \quad (4.108)$$

Ainsi, le meilleur filtrage d'un processus aléatoire normal du mélange additif avec un processus aléatoire normal (dépendant en général du premier) est réalisé par un filtre linéaire optimal dont la fonction de transition impulsionnelle est donnée par l'équation (4.12). Ce résultat peut être généralisé au cas des processus aléatoires normaux non stationnaires.

Le fait que le meilleur filtrage des processus aléatoires normaux soit linéaire est dû à ce que la surface de régression pour un ensemble des variables aléatoires normales coïncide avec le plan de la régression quadratique moyenne (cf. annexe XI).

4.4.5. Interprétation des filtres non linéaires. Exposée dans ce paragraphe la méthode du filtrage non linéaire, optimal du point de vue du critère de minimum du carré moyen de l'erreur, est basée sur l'utilisation de l'approximation d'une fonctionnelle continue par des suites de la forme (4.81). Un filtre de degré n est caractérisé par la suite de noyaux $K_m(u_1, \ldots, u_m)$, $m = 2, 3, \ldots, n$ satisfaisant aux équations intégrales (4.103). En résolvant ces équations on peut en principe, d'après une réalisation donnée $x(t)$ d'un mélange additif des processus aléatoires $\xi(t)$ et $\eta(t)$, formuler conformément à (4.97) l'estimation $\hat{\xi}(t)$ d'un de ces processus pour laquelle la moyenne $m_1\{[\hat{\xi}(t) - \xi(t)]^2\}$ est minimale dans la classe des filtres non linéaires de degré n. La réalisation pratique d'un filtre d'après une suite donnée de noyaux K_n est liée soit à un algorithme compliqué de calcul, soit à une interprétation adéquate des noyaux. L'une de ces interprétations est basée sur le développement d'une fonction de plusieurs variables en séries multiples suivant des polynômes orthogonaux.

Soit $\{Q_n(x)\}$ un ensemble des polynômes orthonormés dont la fonction de pondération est $\varphi(x)$. On peut ainsi, dans de nombreux cas, développer le noyau $K_n(u_1, \ldots, u_n)$ en série multiple d'ordre n, soit:

$$K_n(u_1, \ldots, u_n) = \varphi(u_1) \ldots \varphi(u_n) \times$$
$$\times \sum_{m_1=0}^{\infty} \ldots \sum_{m_n=0}^{\infty} a_{m_1 \ldots m_n} Q_{m_1}(u_1) \ldots Q_{m_n}(u_n), \quad n \geqslant 1. \qquad (4.109)$$

Comme les polynômes sont orthonormés, les coefficients de la série (4.109) sont

$$a_{m_1 \ldots m_n} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} K_n(u_1, \ldots, u_n) Q_{m_1}(u_1) \ldots$$
$$\ldots Q_{m_n}(u_n) \varphi(u_1) \ldots \varphi(u_n) \, du_1 \ldots du_n. \qquad (4.109')$$

Portons (4.109) dans (4.81) et introduisons la notation

$$x_k(t) = \int_{-}^{\infty} Q_k(u) \varphi(u) x(t-u) \, du,$$

il vient

$$y_n[x(t)] = K_0 + \sum_{k=1}^{n} \sum_{m_1=0}^{\infty} \ldots \sum_{m_k=0}^{\infty} a_{m_1 \ldots m_k} x_{m_1}(t) \ldots x_{m_k}(t). \quad (4.110)$$

L'expression (4.110) permet d'interpréter de la manière suivante l'approximation d'une fonctionnelle non linéaire. Soit un ensemble des filtres linéaires dont les fonctions de transition impulsionnelles sont

$$h_k(u) = \varphi(u)\, Q_k(u). \qquad (4.111)$$

Les processus de sortie de ces filtres sont multipliés de toutes les manières possibles, les produits, pris avec des coefficients (4.109′) tenant compte de la spécificité de la fonctionnelle, sont sommés.

Pour des filtres physiquement réalisables la condition $h_k(u) \equiv 0$ pour $u < 0$ se trouve remplie si l'on utilise l'ensemble des polynô-

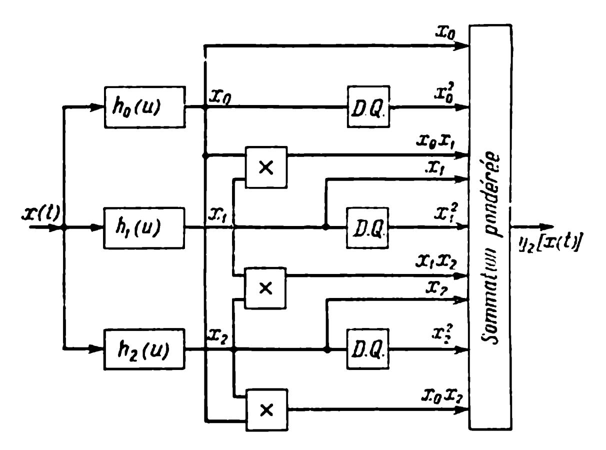


Fig. 4.3. Schéma d'un filtre non linéaire du second degré

mes de Laguerre (cf. annexe IV). Dans ce cas les fonctions de transition impulsionnelles (4.111) appartiennent à un ensemble des fonctions orthogonales de Laguerre

$$h_n(u) = \begin{cases} \dfrac{1}{n!}\dfrac{d^n}{du^n}(u^n e^{-u}), & u \geqslant 0, \\ 0, & u < 0. \end{cases} \qquad (4.112)$$

Donnons à titre d'exemple l'expression approchée de la caractéristique d'un filtre non linéaire du second degré, lorsqu'on utilise l'ensemble des filtres linéaires de degrés zéro, premier et second. En vertu de (4.110) on obtient

$$y_2[x(t)] \approx K_0 + a_0x_0 + a_1x_1 + a_2x_2 + a_{00}x_0^2 + (a_{01} + a_{10})x_0x_1 + $$
$$+ (a_{02} + a_{20})\, x_0x_2 + a_{11}x_1^2 + (a_{12} + a_{21})\, x_1x_2 + a_{22}x_2^2, \qquad (4.113)$$

où

$$x_0 = x_0(t) = \int_0^\infty e^{-u}x(t-u)\, du\,; \qquad (4.114)$$

$$x_1 = x_1(t) = \int_0^\infty (1-u)\, e^{-u} x(t-u)\, du\,; \tag{4.114'}$$

$$x_2 = x_2(t) = \int_0^\infty \left(1 - 2u + \frac{u^2}{2}\right) e^{-u} x(t-u)\, du \tag{4.114''}$$

et

$$a_m = \int_{-\infty}^{\infty} K_1(u)\, h_m(u)\, du, \quad m = 0,\ 1,\ 2, \tag{4.115}$$

$$a_{mn} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(u, v)\, h_m(u)\, h_n(v)\, du\, dv, \quad m, n = 0,\ 1,\ 2. \tag{4.115'}$$

Le dispositif permettant de réaliser un filtre non linéaire du second degré, construit comme indiqué ci-dessus, est schématisé par la figure 4.3.

Problèmes

4.1. Soit $\zeta(t)$ un processus aléatoire, stationnaire au sens général, obtenu à la sortie d'un système linéaire de paramètres aléatoires lorsqu'on applique à l'entrée un processus aléatoire $\xi(t)$ stationnaire au sens général, de spectre énergétique $F_\xi(\omega)$. Montrer que l'estimation linéaire de $\zeta(t)$, optimale vis-à-vis du critère de minimum de l'erreur quadratique, faite sur une réalisation du processus $\xi(t)$, peut être obtenue en faisant passer cette réalisation à travers un filtre dont la fonction de transfert est égale à la moyenne $k_s(i\omega)$ de la fonction de transfert du système linéaire mentionné. Démontrer que l'erreur quadratique moyenne est dans ce cas égale à

$$\varepsilon^2_{\min} = \frac{1}{4\pi} \int_{-\infty}^{\infty} F_\xi(\omega)\, [B_s^{\check{}}(0\ \ \omega) - |k_s(i\omega)|^2]\, d\omega, \tag{1}$$

où $B_s(\tau, \omega)$ est la fonction de corrélation du système linéaire.

4.2. Supposons que dans la position du problème du § 4.2.3 on renonce à la condition de réalisation physique et que l'estimation du processus (4.36) s'effectue d'après une réalisation $x(t)$ observée pour toutes les valeurs réelles de t. Supposons également que les processus $\xi(t)$ et $\eta(t)$ soient stationnaires au sens général, la fonction $g(t, \tau)$ étant telle que $g(t, \tau) = g(t-\tau)$ et $G(i\omega)$ la transformée de Fourier de la fonction $g(t)$. Démontrer que la fonction de transfert du filtre optimal donnant l'estimation de $\zeta(t)$ d'après la réalisation du processus somme $\xi(t) + \eta(t)$ est de la forme

$$k^*(i\omega) = G(i\omega)\, \frac{F_{x\xi}(\omega)}{F_x(\omega)}, \tag{2}$$

et l'erreur quadratique minimale est

$$\varepsilon^2_{\min} = \frac{1}{2\pi} \int_0^\infty [\, |G(i\omega)|^2 F_\xi(\omega) - |k^*(i\omega)|^2 F_x(\omega)]\, d\omega =$$

$$= \frac{1}{2\pi} \int_0^\infty |G(i\omega)|^2 \frac{F_\xi(\omega) F_x(\omega) - |F_{x\xi}(\omega)|^2}{F_x(\omega)} d\omega. \quad (3)$$

4.3. Soit $x(t)$ une réalisation du processus aléatoire somme d'un bruit blanc d'intensité N_0 et du processus aléatoire indépendant $\xi(t)$ obtenu à la sortie d'un circuit intégrateur RC lorsqu'on applique à son entrée un bruit blanc d'intensité N_s. Montrer que la meilleure, vis-à-vis du critère de minimum du carré moyen de l'erreur, estimation de la valeur du processus $\xi(t)$, extrapolée sur le temps t_0, faite sur la réalisation $x(t)$, est de la forme

$$\hat{\xi}(t+t_0) = A_0 \int_0^\infty x(t-u)\, e^{-\alpha\sqrt{1+s^2}u}\, du, \quad t_0 > 0, \quad (4)$$

où

$$A_0 = \frac{\alpha s^2}{1+\sqrt{1+s^2}} e^{-\alpha t_0}; \quad s^2 = \frac{N_s}{N_0}, \quad \alpha = \frac{1}{RC}. \quad (4')$$

Trouver l'expression suivante de la valeur minimale du carré moyen de l'erreur d'extrapolation:

$$\varepsilon^2_{\min} = \frac{\alpha N_s}{4}\left[1 - \frac{s^2}{(1+\sqrt{1+s^2})^2} e^{-2\alpha t_0}\right], \quad t_0 > 0. \quad (5)$$

Obtenir à partir de (4) et (5) à la limite pour $N_0 \to 0$ ($s^2 \to \infty$) les formules suivantes correspondant à l'extrapolation pure:

$$\hat{\xi}(t+t_0) = x(t)\, e^{-\alpha t_0}, \quad t_0 > 0, \quad (6)$$

$$\varepsilon^2_{\min} = \frac{\alpha N_s}{4}(1 - e^{-2\alpha t_0}). \quad (7)$$

Noter que le filtrage à l'instant t s'obtient à partir des formules ci-dessus pour $t_0 = 0$.

4.4. Pour les conditions du problème 4.3 montrer que, pour $t_0 < 0$, la meilleure estimation de la valeur interpolée du processus $\xi(t)$ faite sur la réalisation $x(t)$ est:

$$\hat{\xi}(t-|t_0|) = \frac{\alpha s^2}{\sqrt{1+s^2}} \int_0^\infty x(t-u)\left[e^{-\alpha\sqrt{1+s^2}(u-|t_0|)} - \right.$$

$$\left. - 2\frac{1-\sqrt{1+s^2}}{1+\sqrt{1+s^2}} e^{-\alpha\sqrt{1+s^2}(u+|t_0|)}\right] du. \quad (8)$$

En utilisant (4.31') trouver dans ce cas l'expression de l'erreur quadratique moyenne minimale d'interpolation. Vérifier que pour $N_0 \to 0$ on a

$$\hat{\xi}(t - |t_0|) = x(t - |t_0|),$$

c'est-à-dire que l'estimation se réduit à un retard de $|t_0|$.

4.5. Montrer que le filtre adapté pour un signal continu $s(t) \equiv a$ est un intégrateur parfait de fonction de transition impulsionnelle $h^*(t) = au(t)$, où $u(t)$ est un échelon unité.

4.6. Montrer que pour un signal impulsionnel sinusoïdal

$$s(t) = a \sin \omega_0 t, \quad 0 < t < T, \quad \omega_0 T = (2n+1)\pi, \quad (9)$$

la fonction de transition impulsionnelle du filtre adapté est

$$h^*(u) = a \sin \omega_0 u, \quad u > 0. \quad (10)$$

Chapitre 5

DÉTECTION DES SIGNAUX NOYÉS DANS LES BRUITS

5.1. GÉNÉRALITÉS

Dans la radio-électricité (télécommunications, radar, télécommandes) divers dispositifs doivent résoudre des problèmes liés à l'émission et à la réception de signaux noyés dans les bruits de nature différente. La présence des bruits fait que l'information utile se trouve partiellement ou entièrement perdue. Ceci fait apparaître de nombreux problèmes, en particulier celui de la conception (synthèse) rationnelle de systèmes, consistant à trouver une procédure, optimale vis-à-vis d'un certain critère de qualité, de traitement du processus observé.

Deux types de problèmes peuvent se présenter. Le premier est la *détection* d'un signal en présence de bruits ; il s'agit de décider, d'après les résultats du traitement du processus observé, pouvant être soit un bruit, soit une combinaison d'un signal utile contenant une information codée et d'un bruit, si le signal reçu contient ou non un signal utile. Le second problème, plus général, est dans la sélection d'un signal (ou de plusieurs) dans un groupe de signaux noyés dans les bruits (*discrimination* des signaux).

Si a priori on sait que le signal utile est présent dans le processus observé, le problème se pose de faire *ressortir* ce signal, noyé dans le bruit, en mesurant un certain nombre de paramètres informationnels du signal ou en le décrivant comme une fonction du temps.

Le signal aussi bien que le bruit le déformant sont en général des processus aléatoires. De plus, notre connaissance du processus est basée sur des échantillons de taille finie sous la forme soit d'un ensemble de nombres (échantillon discret), soit d'une ou de plusieurs réalisations tronquées du processus observé (échantillon continu). Ceci explique pourquoi on résout le problème de traitement des processus observés pour en extraire l'information utile par une *méthode de statistique*. Il est déjà de tradition dans la radiotechnique moderne de considérer la réception des signaux comme un problème de statistique.

Dans les chapitres précédents nous avons jeté les bases mathématiques permettant de résoudre les problèmes de détection et de discrimination des signaux noyés dans les bruits. En effet, du point de vue formel, la détection est une vérification des hypothèses statistiques concernant les caractéristiques d'une variable ou d'un processus aléatoires (chapitres 1 et 3), et la discrimination, une estimation des paramètres, ou un filtrage (chapitres 2, 3, 4). Tout comme la véri-

fication des hypothèses et la recherche d'une estimation, les deux problèmes de radiotechnique mentionnés peuvent être étudiés par la théorie des décisions statistiques [6], [15].

Dans ce chapitre et dans les chapitres suivants nous allons étudier en détail l'application des méthodes statistiques aux systèmes radiotechniques. Ces systèmes ont ceci de particulier qu'ils transmettent l'information par ondes électromagnétiques modulées à haute fréquence, par suite, les processus à l'entrée des dispositifs de réception sont *à bande étroite* (en ce sens que la largeur de bande du spectre du processus est bien inférieure à la fréquence centrale du spectre, cf. tome I, § 4.2.4). C'est pourquoi, en plus des processus à spectre arbitraire, nous allons étudier séparément et plus en détail les processus à bande étroite dont les spectres se trouvent essentiellement concentrés dans la bande de fréquences $\omega_0 \pm \frac{\Delta_s}{2}$ $(\Delta_s \ll \omega_0)$. Nous avons montré au § 6.2 du premier tome que ces processus peuvent s'écrire comme suit:

$$x(t) = r(t) \cos [\omega_0 t + \vartheta(t)], \qquad (5.1)$$

où $r(t)$, $\vartheta(t)$ sont l'enveloppe et la phase lentement variables, dont les spectres sont pratiquement limités à la bande de fréquences $(0, \frac{\Delta_s}{2})$. C'est pourquoi, au lieu d'un échantillon composé de valeurs instantanées $(x_1, \ldots, x_N)$ du processus observé, on peut utiliser un échantillon formé par les valeurs de l'enveloppe et de la phase du processus, soit $(r_1, \vartheta_1, r_2, \vartheta_2, \ldots, r_N, \vartheta_N)$.

Parfois il est désirable avant tout traitement du signal reçu d'en séparer l'enveloppe et la phase (c'est-à-dire de procéder à une *détection*), puis d'appliquer une procédure optimale de traitement de l'échantillon $(r_1, \ldots, r_N)$ (méthode d'amplitude) ou $(\vartheta_1, \ldots, \vartheta_N)$ (méthode de phase). C'est pourquoi à côté des méthodes optimales de traitement des processus arrivant à l'entrée des dispositifs de réception (traitement absolument optimal *prédétecteur*) nous allons étudier les méthodes de traitement optimal *postdétecteur*. Il est évident que les méthodes optimales d'amplitude ou de phase ne peuvent pas être meilleures que les procédures optimales absolues, car la détection de l'enveloppe et de la phase est en général inéluctablement liée à une perte d'information.

Dans ce qui suit nous nous bornerons au cas où les bruits sont additifs et statistiquement indépendants du signal et représentent un processus aléatoire normal stationnaire, de moyenne nulle et de spectre énergétique arbitraire. La structure du signal utile sera spécialement mentionnée dans chaque cas particulier.

Dans le présent chapitre nous étudierons les problèmes de détection, et dans le suivant les problèmes de sélection des signaux.

5.2. ALGORITHMES OPTIMAUX DE DÉTECTION D'UN SIGNAL DANS UN BRUIT NORMAL ADDITIF

5.2.1. Signal déterministe. Nous allons commencer par le cas simple de la détection d'un signal déterministe *) $s(t)$ dans un bruit blanc additif normal. Il s'agit de vérifier l'hypothèse simple (H_0) que le processus observé est stationnaire, normal, de moyenne nulle, contre l'alternative simple (H_1) que ce processus est également normal, mais que sa moyenne varie suivant une loi connue $s(t)$.

Pour les valeurs échantillonnées nous allons prendre (cf. § 3.4) les coordonnées non corrélées

$$x_k = \sqrt{\lambda_k} \int_{-T}^{T} x(t)\, \varphi_k(t)\, dt, \tag{5.2}$$

où $x(t)$ est une réalisation du processus observée sur l'intervalle d'observation $(-T, T)$; λ_k et $\varphi_k(t)$ étant les nombres caractéristiques et les fonctions propres de l'équation intégrale

$$\varphi(t) = \lambda \int_{-T}^{T} B(y-t)\, \varphi(y)\, dy, \quad |t| \leqslant T, \tag{5.3}$$

où $B(\tau)$ est la fonction de corrélation donnée du bruit **).

En se limitant aux N premières coordonnées et compte tenu de (3.45) et (3.47) il est facile de voir que le logarithme du rapport de vraisemblance pour un échantillon $x_1, \ldots, x_N$ de taille N peut s'écrire comme suit:

$$\ln l(x_1, \ldots, x_N) = \sum_{i=1}^{N} x_i s_i - \frac{1}{2} \sum_{i=1}^{N} s_i^2, \tag{5.4}$$

$$s_i = \sqrt{\lambda_i} \int_{-T}^{T} s(t)\, \varphi_i(t)\, dt, \quad i = 1, \ldots, N. \tag{5.5}$$

En utilisant la théorie exposée dans le premier chapitre et compte tenu de (5.4) nous pouvons formuler le test permettant de juger de la présence ou de l'absence du signal $s(t)$ d'après les coordonnées observées $x_1, \ldots, x_N$ dans un mélange additif avec un bruit. On adopte la décision (γ_1) sur la présence d'un signal, si pour l'échantillon observé on a

$$\sum_{i=1}^{N} x_i s_i \geqslant \ln c + \frac{1}{2} \sum_{i=1}^{N} s_i^2, \tag{5.6}$$

*) C'est ce qu'on appelle la détection *cohérente*.

**) Les résultats exposés dans le chapitre 3 permettent de généraliser facilement la solution du problème étudié au cas d'un bruit normal non stationnaire de moyenne nulle et dont la fonction de corrélation est $B(t, y)$.

et l'on adopte la décision (γ_0) sur l'absence de signal, si c'est l'inégalité inverse de (5.6) qui est vérifiée.

Ainsi, l'algorithme de la détection se réduit au calcul de la somme $\sum_{i=1}^{N} x_i s_i$ et à sa comparaison avec le seuil

$$K_N = \ln c + \frac{1}{2} \sum_{i=1}^{N} s_i^2 \tag{5.7}$$

indépendant de l'échantillon observé.

Cet algorithme est *optimal* vis-à-vis de tous les critères de qualité envisagés dans le premier chapitre (sauf ceux de l'échantillonnage progressif) : de Bayes (y compris les cas particuliers du maximum de la probabilité a posteriori et du maximum de vraisemblance), critère de Neumann-Pearson et critère du minimax. Seule la constante c (cf. table 1, page 35), c'est-à-dire la grandeur du seuil K_N établi à l'avance conformément à (5.7), dépend du choix du critère de qualité.

Tout comme dans la théorie générale de vérification d'une hypothèse simple contre une alternative simple, lors de la détection d'un signal noyé dans des bruits, l'utilisation d'une stratégie établie à l'avance peut entraîner des erreurs de deux espèces. L'erreur de première espèce est appelée *fausse alarme* : la présence du signal est annoncée alors qu'en réalité il est absent. L'erreur de seconde espèce est appelée *perte du signal* : on annonce l'absence du signal alors qu'en réalité il est présent dans le processus reçu, mais il est masqué par le bruit.

Soient G_1 le domaine de l'espace à N dimensions des échantillons satisfaisant à l'inégalité (5.6) et G_0 la partie restante de l'espace. Les probabilités de fausse alarme α et de perte du signal β s'écrivent comme suit [cf. (1.11) et (1.12)] :

$$\alpha = P\{\gamma_1 \mid H_0\} = \int \cdots \int_{G_1} W_N(x_1, \ldots, x_N \mid 0)\, dx_1 \ldots dx_N, \tag{5.8}$$

$$\beta = P\{\gamma_0 \mid H_1\} = \int \cdots \int_{G_0} W_N(x_1, \ldots, x_N \mid s)\, dx_1 \ldots dx_N. \tag{5.9}$$

La grandeur

$$1 - \beta = P\{\gamma_1 \mid H_1\} \tag{5.10}$$

est appelée probabilité de *détection correcte*.

En utilisant la méthode générale exposée au § 1.2.5, on peut éviter le calcul des intégrales multiples dans (5.8) et (5.9) et les remplacer par des intégrales simples. Les grandeurs α et β s'écrivent

alors comme suit :

$$\alpha = P\{\ln l(x_1, \ldots, x_N) \geqslant \ln c \mid H_0\} = P\left\{\sum_{i=1}^{N} x_i s_i \geqslant K_N \mid H_0\right\}, \quad (5.11)$$

$$\beta = P\{\ln l(x_1, \ldots, x_N) < \ln c \mid H_1\} = P\left\{\sum_{i=1}^{N} x_i s_i < K_N \mid H_1\right\}. \quad (5.12)$$

En vertu de (5.2), les grandeurs x_i sont des variables aléatoires normales indépendantes, de moyennes nulles dans le cas de l'hypothèse H_0 et de moyennes égales à s_i dans le cas de l'hypothèse H_1, la variance dans *tous les cas* étant

$$M_2\{x_i\} = \lambda_k \int_{-T}^{T}\int_{-T}^{T} B(t-y)\,\varphi_i(t)\,\varphi_i(y)\,dt\,dy = \int_{-T}^{T} \varphi_i^2(t)\,dt = 1.$$

La somme des variables aléatoires normales indépendantes $\sum_{i=1}^{N} x_i s_i$ est également une variable aléatoire normale de moyenne nulle dans le cas de l'hypothèse H_0 et de moyenne égale à

$$d_N^2 = \sum_{i=1}^{N} s_i^2, \quad (5.13)$$

dans le cas de l'hypothèse H_1, la variance étant dans les deux cas égale à d_N^2.

On obtient alors à partir de (5.11) et (5.12)

$$\alpha = \frac{1}{\sqrt{2\pi d_N^2}} \int_{K_N}^{\infty} e^{-\frac{z^2}{2d_N^2}}\,dz = 1 - F\left(\frac{K_N}{d_N}\right),$$

$$\beta = \frac{1}{\sqrt{2\pi d_N^2}} \int_{-\infty}^{K_N} e^{-\frac{(z-d_N^2)^2}{2d_N^2}}\,dz = F\left(\frac{K_N - d_N^2}{d_N}\right).$$

Compte tenu de (5.7) on peut écrire comme suit l'expression des probabilités conditionnelles de fausse alarme et de perte du signal :

$$\alpha = 1 - F\left(\frac{\ln c}{d_N} + \frac{1}{2} d_N\right), \quad (5.14)$$

$$\beta = F\left(\frac{\ln c}{d_N} - \frac{1}{2} d_N\right). \quad (5.15)$$

Pour le critère du maximum de vraisemblance (« observateur parfait ») la grandeur c est égale à l'unité. Il découle de (5.14) et (5.15) que les probabilités conditionnelles sont égales entre elles et égales à

$$\alpha = \beta = 1 - F\left(\frac{d_N}{2}\right). \quad (5.15')$$

Il faut noter qu'un test analogue à (5.6) s'obtient également dans le cas où l'on prend pour coordonnées du processus non pas les grandeurs indépendantes x_k données par (5.2) mais les valeurs corrélees du processus à N instants, c'est-à-dire

$$x_k = x(t_k), \; k = 1, \ldots, N.$$

En utilisant la représentation matricielle de la répartition multidimensionnelle normale [cf. (2.57), tome I] on peut écrire comme suit le logarithme du rapport de vraisemblance d'après l'échantillon *corrélé* $x_1, \ldots, x_N$:

$$\ln l(\mathrm{X}) = \mathrm{s'M^{-1}X} - \frac{1}{2}\mathrm{s'M^{-1}s}, \tag{5.16}$$

où X et s sont les vecteurs colonnes dont les éléments sont respectivement $x(t_k)$ et $s(t_k)$ ($k = 1, \ldots, N$), et M^{-1} est la matrice inverse de la matrice de corrélation de bruit.

En introduisant la notation

$$\mathrm{s'M^{-1}} = \mathrm{U}, \tag{5.17}$$

on peut écrire le logarithme du rapport de vraisemblance comme la différence des produits scalaires des vecteurs

$$\ln l(x_1, \ldots, x_N) = \mathrm{UX} - \frac{1}{2}\mathrm{Us} = \sum_{k=1}^{N} u_k x_k - \frac{1}{2}\sum_{k=1}^{N} u_k s_k. \tag{5.18}$$

Le test d'hypothèse s'écrit maintenant comme suit:

$$\sum_{k=1}^{N} u_k x_k \geqslant \ln c + \frac{1}{2}\sum_{k=1}^{N} u_k s_k, \tag{5.19}$$

on en conclut que le signal est présent, si pour l'échantillon donné l'inégalité (5.19) est vérifiée, sinon on conclut à l'absence du signal.

Le test (5.19), tout comme le test (5.6), recommande un traitement linéaire des données échantillonnées, c'est-à-dire la sommation pondérée et la comparaison des résultats de la sommation avec un seuil établi à l'avance dépendant des caractéristiques connues a priori du signal et du critère de qualité. Les probabilités conditionnelles de fausse alarme et de perte du signal sont données par les formules (5.14) et (5.15), où la grandeur d_N est calculée non pas à l'aide de (5.13) mais par la formule suivante

$$d_N^2 = \sum_{k=1}^{N} u_k s_k = \sum_{k=1}^{N}\sum_{j=1}^{N} s_k s_j k_{ij}, \tag{5.20}$$

où k_{ij} sont les éléments de la matrice M^{-1}.

Les tests (5.6) et (5.19) diffèrent en ce que, dans le premier cas, on utilise les coordonnées indépendantes du processus analysé, obte-

nues à partir d'une réalisation observée sur un intervalle donné par intégration pondérée, les poids étant donnés par la solution de l'équation intégrale (5.3), et dans le second cas, les valeurs échantillonnées corrélées, dont on prend la somme pondérée, les poids étant donnés par les composantes du produit du vecteur ligne par la matrice de corrélation de bruit inverse. Dans le premier cas le plus difficile est de trouver la solution de l'équation intégrale (5.3) et dans le second, d'invertir la matrice de corrélation **M**. Notons que pour une même taille N de l'échantillon les caractéristiques pratiques de la détection d'après les algorithmes (5.6) et (5.19) sont en général différentes.

Enfin, considérons l'algorithme optimal de détection dans le cas où les résultats des observations, donnés par une réalisation continue, ne peuvent pas être échantillonnés et les résultats sont traités sous forme analogique et non pas sous forme numérique.

Nous allons utiliser les résultats du § 3.5 en posant $s_1(t) \equiv s(t)$ et $s_0(t) \equiv 0$. On obtient alors à partir de (3.81) le test suivant: le signal est présent si pour la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$ on a

$$\int_{-T}^{T} V(t)\, x(t)\, dt \geqslant \ln c + \frac{1}{2} \int_{-T}^{T} V(t)\, s(t)\, dt = K_T, \tag{5.21}$$

et le signal est absent si l'on a l'inégalité inverse de (5.21). Tout comme dans les formules précédentes la grandeur c est donnée par le critère de qualité, et la fonction $V(t)$ est la solution de l'équation intégrale linéaire non homogène

$$\int_{-T}^{T} B(t-u)\, V(u)\, du = s(t), \quad |t| \leqslant T. \tag{5.22}$$

En vertu de (3.87) et (3.88), les probabilités conditionnelles de fausse alarme et de perte du signal sont données par les formules (5.14) et (5.15) après substitution à d_N de la grandeur d_T calculée à l'aide de la formule suivante [cf. (3.86)] :

$$d_T = \left(\int_{-T}^{T} V(t)\, s(t)\, dt \right)^{1/2}. \tag{5.23}$$

On remarque facilement l'analogie des formules (5.6), (5.19), (5.21) et (5.13), (5.20), (5.23).

Notons que lors de la détection d'un signal déterministe dans un bruit blanc normal additif, en vertu de (3.79), la fonction $V(t)$ dans les formules (5.21) à (5.23) est égale à

$$V(t) = \frac{1}{N_0}\, s(t). \tag{5.24}$$

Désignant par E_s l'énergie du signal sur l'intervalle d'observation

$$E_s = \int_{-T}^{T} s^2(t)\,dt, \tag{5.25}$$

on obtient à partir de (5.21) le test suivant de détection d'un signal dans un bruit blanc : le signal est présent si

$$\frac{1}{N_0}\int_{-T}^{T} s(t)\,x(t)\,dt \geqslant \ln c + \frac{E_s}{2N_0}. \tag{5.26}$$

Le paramètre d_T^2 donnant les probabilités conditionnelles de fausse alarme et de perte du signal est dans ce cas égal à [cf. (5.23) à (5.25)]

$$d_T^2 = \frac{E_s}{N_0}, \tag{5.27}$$

c'est-à-dire au rapport de l'énergie du signal à la densité spectrale du bruit.

En introduisant la puissance du signal $\frac{E_s}{2T}$ et la puissance moyenne du bruit N_0F dans la bande effective F du signal, on peut écrire la formule (5.27) comme suit :

$$d_T^2 = \frac{E_s}{N_0} = 2FTs^2, \tag{5.27'}$$

où s^2 est le rapport de la puissance du signal à la puissance moyenne du bruit.

En éliminant la grandeur $\ln c$ des deux relations (5.14) et (5.15) on obtient la relation analytique suivante entre la probabilité de détection correcte et la probabilité de fausse alarme [comparer avec (1.77')] :

$$x_\alpha - x_{1-\beta} = d, \tag{5.28}$$

où x_α et $x_{1-\beta}$ sont les quantiles de la loi normale, et le paramètre d est donné par l'une des formules (5.13), (5.20) ou (5.23), suivant la méthode de traitement adoptée.

Sur la figure 5.1 on a représenté en trait plein la famille de courbes

$$1 - \beta = f(\alpha;\ d), \tag{5.29}$$

cette famille est appelée *caractéristique de travail de la détection*. La ligne en trait mixte fin coupe les courbes en des points correspondant au critère du maximum de vraisemblance (« observateur parfait »).

Il y a lieu de mentionner une interprétation importante de l'algorithme de détection (5.21). En utilisant les résultats du § 4.3.1 on trouve que le traitement *linéaire* optimal de la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$ dans le but de détecter un signal

déterministe $s(t)$ consiste à faire passer cette réalisation à travers un filtre adapté dont la fonction de transition impulsionnelle est

$$h^*(\tau) = V(T - \tau), \quad |\tau| \leqslant T,$$

où $V(t)$ est la solution de l'équation intégrale (5.22). Comme nous l'avons déjà noté au § 4.3.4, ce filtre peut être un *corrélomètre*. Dans le cas d'une détection dans un bruit blanc additif, le filtre adapté

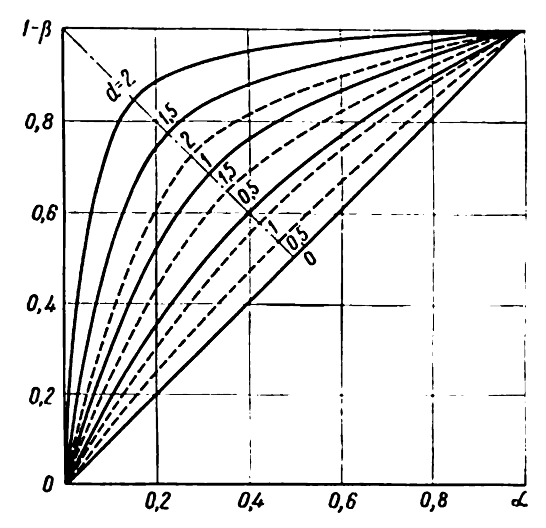


Fig. 5.1. Caractéristiques de la détection des signaux déterministe (trait plein) et quasi déterministe (pointillé)

est plus facile à réaliser, car la fonction de transition impulsionnelle est tout simplement l'image du signal. Le corrélomètre correspondant donne simplement la fonction de corrélation mutuelle temporelle de la réalisation observée et du signal déterministe.

Ainsi, le dispositif optimal de détection d'un signal déterministe noyé dans un bruit blanc additif normal est un filtre adapté suivi d'un dispositif non inertiel à seuil.

5.2.2. Signal quasi déterministe. Considérons maintenant le problème plus compliqué de la détection d'un signal *quasi déterministe* qui est un processus à bande étroite dont l'amplitude $a(t)$ et la phase $\psi_s(t)$ sont modulées suivant une loi donnée et dont la phase initiale φ_0 est aléatoire *), soit:

$$s(t) = a(t) \cos [\omega_0 t - \psi_s(t) + \varphi_0], \tag{5.30}$$

noyé dans un bruit blanc additif normal à bande étroite. Il s'agit de vérifier l'hypothèse simple H_0 selon laquelle le processus observé

*) C'est ce qu'on appelle une *détection non cohérente*.

est stationnaire, normal, de moyenne nulle, contre l'alternative *composée* H_1 que ce processus est également normal, mais de moyenne $s(t)$ qui est l'une des valeurs du continuum de réalisations correspondant aux variations de la phase aléatoire φ_0 sur l'intervalle $(-\pi, \pi)$.

Nous allons utiliser la représentation complexe de la réalisation d'un processus aléatoire à bande étroite (cf. § 6.2.2 du premier tome et annexe VI)

$$x(t) = \operatorname{Re} z(t)\, e^{i\omega_0(t)}, \tag{5.31}$$

où $z(t)$ est l'enveloppe complexe du processus aléatoire à bande étroite, liée à ses enveloppe $r(t)$ et phase $\vartheta(t)$ par la relation

$$z(t) = r(t)\, e^{i\vartheta(t)}, \tag{5.32}$$

de plus

$$z(t) = A(t) + iC(t), \tag{5.32'}$$

où $A(t)$ et $C(t)$ sont les *composantes en quadrature* d'un processus aléatoire à bande étroite. Dans un processus normal ces composantes sont également normales. L'enveloppe complexe du signal quasi déterministe (5.30) désignée par $z_s(t)$ est égale à

$$z_s(t) = a(t)\, e^{-i\psi_s(t)}, \tag{5.33}$$

$$s(t) = \operatorname{Re} z_s(t)\, e^{i\omega_0 t} e^{i\varphi_0}. \tag{5.33'}$$

On peut maintenant caractériser la réalisation $x(t)$ sur l'intervalle d'observation $(-T, T)$ par un ensemble des coordonnées non corrélées de l'enveloppe complexe

$$z_k = \sqrt{\lambda_k} \int_{-T}^{T} z(t)\, \overline{\varphi_k(t)}\, dt. \tag{5.34}$$

Ici λ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale *)

$$\varphi(t) = \lambda \int_{-T}^{T} B_z(t-u)\, \varphi(u)\, du, \tag{5.35}$$

où

$$B_z(\tau) = \frac{1}{2} m_1 \{z(t)\, \overline{z(t+\tau)}\} \tag{5.35'}$$

est la fonction de corrélation de l'enveloppe complexe du bruit, dont les parties réelle et imaginaire coïncident avec les fonctions de corrélation et de corrélation mutuelle des composantes en quadrature

$$\operatorname{Re} B_z(\tau) = B_A(\tau) = B_C(\tau), \tag{5.36}$$

$$\operatorname{Im} B_z(\tau) = B_{AC}(\tau) = -B_{CA}(\tau). \tag{5.36'}$$

*) Il est également facile de voir que la fonction $B_z(t, u)$ d'un processus complexe est symétrique et définie positive (cf. § 3.3.1), les nombres caractéristiques λ_k étant réels et positifs.

A partir de (5.34) et (5.35) on a

$$m_1\{z_k\bar{z}_m\}=\begin{cases}0, & k\neq m,\\ 2, & k=m.\end{cases} \tag{5.37}$$

En utilisant (5.32), (5.36) et (5.36′) il est facile de voir que

$$m_1\ \{z_k z_m\}\ =0 \tag{5.37′}$$

pour tous k et m.

Soit x_k et y_k les parties réelle et imaginaire de la coordonnée z_k, on a en vertu de (5.37), (5.37′) (cf. § 3.4.5)

$$m_1\ \{x_k y_m\}\ =0 \text{ pour } k \text{ et } m \text{ quelconques,} \tag{5.38}$$

$$m_1\ \{x_k^2\}=m_1\ \{y_k^2\}\ =1, \tag{5.38′}$$

$$m_1\ \{x_k x_m\}\ =m_1\ \{y_k y_m\}\ =0,\ k\neq m. \tag{5.38″}$$

Pour un bruit sans signal (hypothèse H_0) on a $m_1\ \{z\ (t)\}\ =0$ et par conséquent

$$m_1\ \{x_k\ |\ H_0\}\ =m_1\ \{y_k\ |\ H_0\}\ =m_1\ \{z_k\ |\ H_0\}\ =0, \tag{5.39}$$

et pour un bruit avec signal (hypothèse H_1) on a $m_1\ \{z\ (t)\}\ = = z_s\ (t)\ e^{i\varphi_0}$, donc

$$m_1\ \{z_k\ |\ H_1\}\ =s_k e^{i\varphi_0}, \tag{5.39′}$$

$$\left.\begin{aligned} m_1\ \{x_k\ |\ H_1\}\ &=a_k\cos\varphi_0-b_k\sin\varphi_0,\\ m_1\ \{y_k\ |\ H_1\}\ &=a_k\sin\varphi_0+b_k\cos\varphi_0,\end{aligned}\right\} \tag{5.39″}$$

où

$$s_k=a_k+ib_k=\sqrt{\lambda_k}\int_{-T}^{T} z_s(t)\,\overline{\varphi_k(t)}\,dt. \tag{5.39‴}$$

Les composantes de l'enveloppe complexe d'un processus normal à bande étroite étant normales, les grandeurs $x_1,\ y_1,\ \ldots,\ x_N,\ y_N$ forment un ensemble de variables aléatoires indépendantes. L'expression du logarithme du rapport de vraisemblance correspondant à ces $2N$ coordonnées du processus observé peut s'écrire comme suit [cf. (3.71)]:

$$\ln l(x_1,\ y_1,\ \ldots,\ x_N,\ y_N\,|\,\varphi_0)=\frac{1}{2}\sum_{k=1}^{N}[(x_k^2+y_k^2)-$$

$$-(x_k-a_k\cos\varphi_0+b_k\sin\varphi_0)^2-(y_k-a_k\sin\varphi_0-b_k\cos\varphi_0)^2] \tag{5.40}$$

ou sous forme complexe

$$\ln l(z_1,\ \ldots,\ z_N\,|\,\varphi_0)=\sum_{k=1}^{N}\left[\operatorname{Re}(z_k\bar{s}_k e^{-i\varphi_0})-\frac{1}{2}\,|\,s_k\,|^2\right]. \tag{5.41}$$

En passant à la construction du test d'hypothèse (de la stratégie), prenons tout d'abord pour critère de qualité celui de Bayes. Dans ce cas, comme nous l'avons montré au § 1.4.2 [cf. (1.110')] le test optimal est basé sur la comparaison avec un seuil (dépendant de la probabilité a priori de la présence du signal et du coût d'une solution erronée) du rapport *moyen* de vraisemblance. Comme dans le problème envisagé nous avons supposé que le paramètre aléatoire φ_0 soit uniformément réparti, on a à partir de (5.41) (cf. également § 3.5.3)

$$\Lambda(z_1, \ldots, z_N) = \frac{1}{2\pi} \int_0^{2\pi} l(z_1, \ldots, z_N \mid \varphi_0)\, d\varphi_0 =$$

$$= \exp\left\{-\frac{1}{2} \sum_{k=1}^{N} |s_k|^2\right\} \frac{1}{2\pi} \int_0^{2\pi} e^{r_N \cos(\varphi_0 - \psi_N)}\, d\varphi_0,$$

où

$$r_N = \left| \sum_{k=1}^{N} z_k \bar{s}_k \right|; \tag{5.42}$$

$$\psi_N = \operatorname{arc\,tg} \frac{\operatorname{Im} \sum_{k=1}^{N} z_k \bar{s}_k}{\operatorname{Re} \sum_{k=1}^{N} z_k \bar{s}_k}. \tag{5.42'}$$

Utilisons la représentation intégrale connue d'une fonction de Bessel de l'argument imaginaire (cf, par exemple, tome I, § 3.2.2), il vient finalement

$$\Lambda(z_1, \ldots, z_N) = I_0(r_N) \exp\left\{-\frac{1}{2} \sum_{k=1}^{N} |s_k|^2\right\}. \tag{5.43}$$

Si on se limite aux N premières coordonnées $z_1, \ldots, z_N$ de l'enveloppe complexe du processus observé, le test de Bayes permettant de juger de la présence ou de l'absence d'un signal quasi déterministe à bande étroite dans un mélange additif avec un bruit peut se formuler comme suit : on adopte la décision γ_1 (le signal est présent) si

$$\ln I_0(r_N) \geqslant \ln c + \frac{1}{2} \sum_{k=1}^{N} |s_k|^2, \tag{5.44}$$

et la décision γ_0 (le signal est absent) si c'est l'inégalité inverse de (5.44) qui est vérifiée.

L'inégalité (5.44)] [comparer avec (5.6)] peut s'écrire comme suit :

$$r_N \geqslant c_N, \tag{5.44'}$$

où c_N est donné par l'équation transcendante

$$\ln I_0(c_N) = K_N = \ln c + \frac{1}{2} \sum_{k=1}^{N} |s_k|^2. \tag{5.45}$$

Ainsi, l'algorithme optimal (de Bayes) de détection du signal se réduit ici au calcul de la grandeur r_N (c'est-à-dire à un traitement essentiellement *non linéaire* des coordonnées observées) et à la comparaison du résultat obtenu avec un seuil c_N donné par (5.45) et indépendant de l'échantillon.

Les caractéristiques probabilistes de détection dépendent avant tout de la loi de répartition de la variable aléatoire r_N donnée par (5.42). On peut facilement trouver cette loi. En effet, les variables aléatoires z_k sont des variables aléatoires complexes, normales, indépendantes, de moyenne nulle dans le cas de l'hypothèse H_0 et de moyenne égale à $s_k e^{i\varphi_0}$ dans le cas de l'hypothèse H_1. La variance de ces variables pour les deux hypothèses mentionnées est, en vertu de (2.37), égale à

$$M_2\{|z_k|\} = m_1\{|z_k|^2\} = m_1\{z_k \bar{z}_k\} = 2. \tag{5.46}$$

Par conséquent, la variable aléatoire r_N est le module d'une variable aléatoire normale complexe (ou d'un vecteur aléatoire plan) de composantes indépendantes, dont la variance est égale à $2 \sum_{k=1}^{N} |s_k|^2$ et la moyenne est nulle pour l'hypothèse H_0 et égale à $\sum_{k=1}^{N} |s_k|^2 e^{i\varphi_0}$ pour l'hypothèse H_1.

La répartition du module de ce vecteur a été étudiée en détail au § 3.2.2 du premier tome. Conformément au paragraphe cité, la variable aléatoire r_N est répartie suivant la loi de Rayleigh

$$w_1(r) = \frac{r}{d_N^2} e^{-\frac{r^2}{2d_N^2}}, \quad r > 0, \tag{5.47}$$

pour l'hypothèse H_0 et suivant la loi *de Rayleigh généralisée*

$$w_1(r) = \frac{r}{d_N^2} e^{-\frac{r^2 + d_N^4}{2d_N^2}} I_0(r), \quad r > 0, \tag{5.48}$$

pour l'hypothèse H_1.

Dans les formules (5.47) et (5.48) le paramètre d_N est donné par la formule [comparer avec (5.13)]

$$d_N^2 = \sum_{k=1}^{N} |s_k|^2. \tag{5.49}$$

On peut maintenant écrire les expressions donnant la probabilité de fausse alarme ou de perte du signal. Comme la fonction $\ln I_0(r)$ croît d'une manière monotone, en vertu du test optimal (5.44) on a

$$\alpha = P\{\ln I_0(r_N) \geqslant K_N \mid H_0\} = P\{r_N \geqslant c_N \mid H_0\}, \quad (5.50)$$

$$\beta = P\{\ln I_0(r_N) < K_N \mid H_1\} = P\{r_N < c_N \mid H_1\}, \quad (5.51)$$

où c_N est l'abscisse du point d'intersection de la courbe $y = \ln I_0(r)$ et de la droite $y = K_N$ [cf. (5.45)].

En vertu de (5.50) et (5.51) et compte tenu de (5.47) et (5.48) on a

$$\alpha = \int_{c_N}^{\infty} \frac{r}{d_N^2} e^{-\frac{r^2}{2d_N^2}} dr = e^{-\frac{c_N^2}{2d_N^2}}, \quad (5.52)$$

$$\beta = e^{-\frac{d_N^2}{2}} \int_0^{\frac{c_N}{d_N}} x e^{-\frac{x^2}{2}} I_0(xd_N)\, dx. \quad (5.53)$$

Cette dernière intégrale est tabulée.

En plus de la procédure étudiée ci-dessus de détection d'après un échantillon discret, nous allons trouver un algorithme analogue dans le cas où les résultats des observations ne peuvent pas être échantillonnés et sont donnés par une réalisation continue $x(t)$ sur l'intervalle $(-T, T)$.

Compte tenu de (5.35′) et après substitution de $z_s(t)\, e^{i\varphi_0}$ au lieu de $s(t)$ on obtient l'expression suivante de la *fonctionnelle* du rapport de vraisemblance correspondant à l'enveloppe $z(t)$ du processus pour φ donné, observée sur l'intervalle $(-T, T)$:

$$l[z(t) \mid \varphi_0] = \exp\left\{\operatorname{Re} \int_{-T}^{T} V(t;\, \varphi_0)\, \overline{z(t)}\, dt\right\} \times$$
$$\times \exp\left\{-\frac{1}{2} \operatorname{Re} \int_{-T}^{T} V(t;\, \varphi_0)\, \overline{z_s(t)}\, e^{-i\varphi_0} dt\right\}, \quad (5.54)$$

où $V(t;\, \varphi_0)$ est donné par l'équation intégrale linéaire non homogène [cf. (3.77)]

$$\int_{-T}^{T} B_z(t-u)\, V(u;\, \varphi_0)\, du = z_s(t)\, e^{i\varphi_0}, \quad |t| \leqslant T. \quad (5.54')$$

En introduisant la fonction

$$U(t) = V(t;\, \varphi_0)\, e^{-i\varphi_0}, \quad (5.55)$$

on peut localiser le paramètre φ_0 dans l'expression de la fonctionnelle du rapport de vraisemblance. En portant (5.55) dans (5.54) on obtient

pour $U(t)$ l'équation intégrale

$$\int_{-T}^{T} B_z(t-y)\,U(y)\,dy = z_s(t), \quad |t| \leqslant T, \tag{5.56}$$

montrant que la fonction $U(t)$ ne dépend pas de φ_0. Notons que l'équation intégrale *complexe* (5.56) est équivalente à un système de deux équations intégrales réelles pour la partie réelle $u(t)$ et la partie imaginaire $v(t)$ de la fonction $U(t)$:

$$\int_{-T}^{T} [B_A(t-y)\,u(y) - B_{AC}(t-y)\,v(y)]\,dy = a(t)\cos\psi_s(t), \tag{5.56'}$$

$$\int_{-T}^{T} [B_A(t-y)\,v(y) + B_{AC}(t-y)\,u(y)]\,dy = a(t)\sin\psi_s(t), \quad |t| \leqslant T. \tag{5.56''}$$

L'expression (5.54) de la fonctionnelle du rapport de vraisemblance peut, compte tenu de (5.55), s'écrire comme suit:

$$l[z(t)\,|\,\varphi_0] = \exp\left\{\operatorname{Re} e^{i\varphi_0}\int_{-T}^{T} U(t)\,\overline{z(t)}\,dt\right\}\times$$
$$\times\exp\left\{-\frac{1}{2}\operatorname{Re}\int_{-T}^{T} U(t)\,\overline{z_s(t)}\,dt\right\} =$$
$$= \exp\left\{\left|\int_{-T}^{T} U(t)\,\overline{z(t)}\,dt\right|\cos(\varphi_0-\psi_T)\times\right.$$
$$\left.\times\exp\left\{-\frac{1}{2}\int_{-T}^{T} U(t)\,\overline{z_s(t)}\,dt\right\}\right., \tag{5.57}$$

où

$$\psi_T = \operatorname{arc\,tg}\frac{\operatorname{Im}\int_{-T}^{T} U(t)\,\overline{z(t)}\,dt}{\operatorname{Re}\int_{-T}^{T} U(t)\,\overline{z(t)}\,dt}; \tag{5.57'}$$

le symbole Re dans le second facteur de (5.57) a été omis car l'intégrale est réelle et positive

$$\int_{-T}^{T} U(t)\,\overline{z_s(t)}\,dt = \overline{\int_{-T}^{T} \overline{U(t)}\,z_s(t)\,dt} = \overline{\int_{-T}^{T}\int_{-T}^{T} B_z(t-y)\,U(y)\,\overline{U(t)}\,dy\,dt} =$$
$$= m_1\left\{\left|\int_{-T}^{T} \overline{z(t)}\,U(t)\,dt\right|^2\right\} > 0. \tag{5.58}$$

En prenant la moyenne de la fonctionnelle (5.57) sur la phase aléatoire uniformément répartie (sur l'intervalle $|\varphi_0| \leqslant \pi$ on obtient :

$$\Lambda [z(t)] = I_0(r_T)\, e^{-\frac{d_T^2}{2}}, \tag{5.59}$$

on a introduit ici les notations [analogues à (5.42) et (5.49)]

$$r_T = \left| \int_{-T}^{T} U(t)\, \overline{z(t)}\, dt \right|; \tag{5.60}$$

$$d_T^2 = \int_{-T}^{T} U(t)\, \overline{z_s(t)}\, dt. \tag{5.61}$$

Le test optimal (conformément au critère bayesien de risque minimal) peut maintenant se formuler comme suit : le signal est présent si pour la réalisation de l'enveloppe complexe observée sur l'intervalle $(-T, T)$ on a

$$\ln I_0(r_T) \geqslant \ln c + \frac{d_T^2}{2} = K_T \tag{5.62}$$

ou

$$r_T \geqslant c_T, \tag{5.62'}$$

la grandeur c_T étant donnée par l'équation transcendante

$$\ln I_0(c_T) = K_T. \tag{5.63}$$

La variable aléatoire r_T est le module de la variable aléatoire complexe normale $\int_{-T}^{T} U(t)\, \overline{z(t)}\, dt$ dont la moyenne est nulle pour l'hypothèse H_0 et égale à $d_T^2 e^{i\varphi_0}$ pour l'hypothèse H_1. En vertu de (5.58) la variance de cette variable est toujours égale à [cf. (5.35')]

$$M_2 \left\{ \left| \int_{-T}^{T} U(t)\, \overline{z(t)}\, dt \right| \right\} = m_1 \left\{ \left| \int_{-T}^{T} U(t)\, \overline{z(t)}\, dt \right|^2 \right\} = 2d_T^2. \tag{5.64}$$

Ainsi, la variable aléatoire r_T est répartie suivant une loi de Rayleigh de paramètre d_T^2 pour l'hypothèse H_0 et suivant la loi de Rayleigh généralisée de paramètres (d_T^4, d_T^2) pour l'hypothèse H_1. C'est pourquoi les probabilités de fausse alarme et de perte du signal sont données comme dans le cas discret par les formules (5.52), (5.53) en substituant évidemment à d_N, K_N, c_N les grandeurs d_T, K_T, c_T, cette dernière étant donnée par l'équation transcendante (5.63).

Lors de la détection d'un signal quasi déterministe à bande étroite dans un bruit normal additif de densité spectrale N_0, la fonction

$U(t)$ dans les formules (5.60), (5.61) est égale à

$$U(t)=\frac{1}{N_0}z_s(t). \tag{5.65}$$

L'algorithme optimal de détection se réduit alors au calcul de la grandeur *)

$$r_T=\frac{1}{N_0}\left|\int_{-T}^{T} z_s(t)\,\overline{z(t)}\,dt\right| \tag{5.66}$$

et à sa comparaison avec le seuil c_T donné par (5.63), le paramètre d_T^2 étant égal à

$$d_T^2=\frac{1}{N_0}\int_{-T}^{T}|z_s(t)|^2\,dt, \tag{5.66'}$$

c'est-à-dire qu'il coïncide avec le rapport de l'énergie d'un signal à bande étroite à la densité spectrale du bruit [comparer avec (5.27)].

Sur la figure 5.1 on a représenté en pointillé la caractéristique de détection d'un signal quasi déterministe à bande étroite. En la comparant avec les courbes données en trait plein pour les mêmes valeurs du paramètre d_T on peut voir combien elles sont pires par rapport au cas d'un signal déterministe.

Dans le cas de la détection d'un signal harmonique $s(t) = A_0\cos(\omega_0 t+\varphi_0)$ de phase aléatoire, on obtient à partir de (5.66)

$$r_T=\frac{A_0}{N_0}\left[\left(\int_{-T}^{T}A(t)\,dt\right)^2+\left(\int_{-T}^{T}C(t)\,dt\right)^2\right]^{\frac{1}{2}}. \tag{5.67}$$

Comme les composantes en quadrature $A(t)$ et $C(t)$ de la réalisation observée $x(t)$ varient lentement par rapport à $\cos\omega_0 t$ **), on

---

*) Un dispositif optimal de détection peut calculer au lieu de r_T une fonction monotone quelconque $f(r_T)$ et la comparer avec $f(c_T)$ sans changer les valeurs des probabilités d'erreurs. Ainsi, la caractéristique de détection est donnée non pas par le rapport signal/bruit à la *sortie* du dispositif, rapport dépendant de la forme de la fonction $f(r_T)$, mais par le rapport de l'énergie du signal à la densité spectrale du bruit blanc (ou par la grandeur analogue (5.61) quand le spectre du bruit n'est pas uniforme).

**) Par exemple, on a

$$\int_{-T}^{T}x(t)\cos\omega_0 t\,dt=\int_{-T}^{T}A(t)\cos^2\omega_0 t\,dt+\frac{1}{2}\int_{-T}^{T}C(t)\sin 2\omega_0 t\,dt=$$

$$=\frac{1}{2}\int_{-T}^{T}A(t)\,dt+\frac{1}{2}\int_{-T}^{T}[A(t)\cos 2\omega_0 t+C(t)\sin 2\omega_0 t]\approx\frac{1}{2}\int_{-T}^{T}A(t)\,dt.$$

peut écrire (5.67) comme suit

$$r_T = \frac{2A_0}{N_0}\left[\left(\int_{-T}^{T} x(t)\cos\omega_0 t\,dt\right)^2 + \left(\int_{-T}^{T} x(t)\sin\omega_0 t\,dt\right)^2\right]^{\frac{1}{2}}. \qquad (5.67')$$

Le paramètre d_T^2 est ici égal à

$$d_T^2 = \frac{2TA_0^2}{N_0}.$$

Notons qu'en vertu de (3.102) le domaine critique, pour un φ_0 donné, où l'hypothèse H_0 est rejetée (c'est-à-dire qu'on adopte la décision que le signal est présent) se détermine d'après le critère de Neumann-Pearson par l'inégalité suivante:

$$\mathrm{Re}\int_{-T}^{T} V(t;\ \varphi_0)\,\overline{z(t)}\,dt = r_T\cos(\varphi_0 - \psi_T) \geqslant K(\varphi_0), \qquad (5.68)$$

où $K(\varphi_0)$ est donné par une certaine probabilité α de fausse alarme. La relation (5.68) montre que la stratégie uniformément la plus puissante dans le cas envisagé *n'existe pas*. On peut modifier ce critère en utilisant le test (5.68) centré sur φ_0, c'est-à-dire le test (5.62′) où c_T est donné par la formule [cf. (5.52)]

$$c_T = d_T\sqrt{2\ln\frac{1}{\alpha}}. \qquad (5.69)$$

L'algorithme optimal de détection (5.62′) peut être interprété assez simplement en passant des grandeurs complexes aux grandeurs réelles. On peut alors écrire (5.62′) comme suit

$$(\mathrm{Re}\,r_T)^2 + (\mathrm{Im}\,r_T)^2 \geqslant c_T^2, \qquad (5.70)$$

où

$$\mathrm{Re}\,r_T = \int_{-T}^{T} [u(t)A(t) + v(t)C(t)]\,dt; \qquad (5.71)$$

$$\mathrm{Im}\,r_T = \int_{-T}^{T} [v(t)A(t) - u(t)C(t)]\,dt. \qquad (5.71')$$

Soient $h_c^*(\tau)$ et $h_s^*(\tau)$ les fonctions de transition impulsionnelles des filtres adaptés aux composantes en quadrature $a(t)\cos\psi_s(t)$ et $a(t)\sin\psi_s(t)$ du signal. Ces fonctions sont données par les solutions des systèmes d'équations intégrales linéaires (5.56′) et (5.56″):

$$h_c^*(\tau) = u(T-\tau), \quad |\tau| \leqslant T, \qquad (5.72)$$

$$h_s^*(\tau) - v(T-\tau), \quad |\tau| \leqslant T. \qquad (5.72')$$

Le traitement *non linéaire* optimal de la réalisation $x(t)$ observée sur

l'intervalle ($-T$, T) pour la détection d'un signal quasi déterministe à bande étroite dans un bruit blanc additif consiste à (fig. 5.2): 1) faire passer les composantes en quadrature du processus observé à

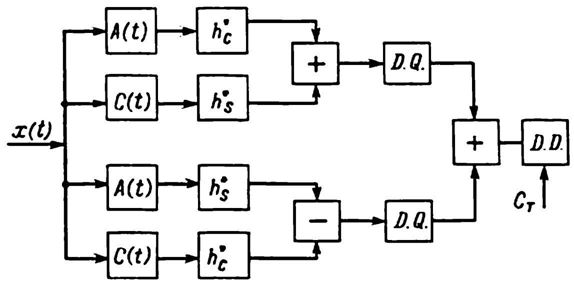


Fig. 5.2. Schéma d'un dispositif optimal de détection d'un signal quasi déterministe à bande étroite

travers deux groupes de filtres adaptés aux composantes en quadrature du signal; 2) former la somme et la différence des valeurs du signal de sortie dans chaque groupe de filtres; 3) effectuer une détection quadratique à deux alternances de la somme et de la différence des valeurs du signal; 4) prendre la somme des grandeurs détectées.

Si le signal est non modulé en phase, on a $h_s^*(\tau) \equiv 0$ et dans chacun des groupes mentionnés de filtres il ne reste plus qu'un seul filtre adapté à l'enveloppe du signal $a(t)$.

Ainsi, le dispositif optimal de détection d'un signal quasi déterministe à bande étroite noyé dans un bruit additif normal se compose d'un système de filtres, d'additionneurs et de détecteurs quadratiques suivis d'un élément non inertiel à seuil. En pratique on utilise un dispositif optimal équivalent plus simple se composant d'un filtre adapté au signal et d'un détecteur linéaire séparant l'enveloppe.

5.2.3. Signal stochastique. Supposons que le signal utile, tout comme le bruit additif indépendant du signal, soient des processus aléatoires normaux stationnaires de moyenne nulle, ne différant que par leurs spectres (c'est-à-dire par leurs fonctions de corrélation). Soient $B_s(\tau)$ et $B_b(\tau)$ les fonctions de corrélation données du signal et du bruit. Pour détecter ce signal stochastique il y a lieu de vérifier l'hypothèse H_0 selon laquelle la fonction de corrélation du processus aléatoire normal stationnaire est égale à $B_b(\tau)$ contre l'alternative H_1 que la fonction de corrélation du processus est $B_s(\tau) + B_b(\tau)$.

Nous allons prendre, conformément au § 3.5.4, pour coordonnées observées les coordonnées non corrélées

$$x_k = \sqrt{\lambda_k} \int_{-T}^{T} x(t)\, \varphi_k(t)\, dt, \tag{5.73}$$

où $x(t)$ est la réalisation du processus sur l'intervalle $(-T, T)$ et λ_k et $\varphi_k(t)$ les nombres caractéristiques et les fonctions propres (non normées) de l'équation intégrale linéaire [cf. (3.120)]

$$\int_{-T}^{T} B_s(t-y)\,\varphi(y)\,dy = (\lambda-1)\int_{-T}^{T} B_b(t-y)\,\varphi(y)\,dy, \quad |t| \leqslant T, \tag{5.74}$$

la normalisation des fonctions propres étant effectuée conformément à l'égalité

$$\lambda_k \int_{-T}^{T}\int_{-T}^{T} B_b(t-y)\,\varphi_k(t)\,\varphi_m(y)\,dt\,dy = \begin{cases} 1, & k=m, \\ 0, & k \neq m. \end{cases} \tag{5.74'}$$

En vertu de (3.122) le logarithme du rapport de vraisemblance pour l'échantillon $x_1, \ldots, x_N$ de taille finie N est

$$\ln l(x_1, \ldots, x_N) = \frac{1}{2}\sum_{k=1}^{N} \frac{\lambda_k - 1}{\lambda_k} x_k^2 - \frac{1}{2}\sum_{k=1}^{N} \ln \lambda_k. \tag{5.75}$$

Le test se formule maintenant comme suit : le signal est présent si

$$\sum_{k=1}^{N} \frac{\lambda_k - 1}{\lambda_k} x_k^2 \geqslant 2\ln c + \sum_{k=1}^{N} \ln \lambda_k, \tag{5.76}$$

et est absent si c'est l'inégalité inverse de (5.76) qui est vérifiée. Ainsi, l'algorithme de détection d'un signal se réduit au calcul de la somme pondérée des carrés des données échantillonnées (coordonnées non corrélées) et à sa comparaison avec un seuil fixé à l'avance

$$K_N = 2\ln c + \sum_{k=1}^{N} \ln \lambda_k. \tag{5.76'}$$

Cet algorithme est *optimal* vis-à-vis de tous les critères (sauf les critères d'échantillonnage progressif) mentionnés dans le premier chapitre. Le choix du critère n'influe que sur la valeur du seuil (constante c).

Les probabilités conditionnelles de fausse alarme et de perte d'un signal sont données par les répartitions (3.126), (3.126') :

$$\alpha = \frac{1}{2\pi}\int_{\ln c}^{\infty}\int_{-\infty}^{\infty} \left[\prod_{k=1}^{N}\left(1 - \frac{iv}{1+v_k}\right)\right]^{-\frac{1}{2}} \times$$
$$\times \exp\left\{-\frac{iv}{2}\ln\prod_{k=1}^{N}\left(1+\frac{1}{v_k}\right)\right\} e^{ivx}\,dv\,dx, \tag{5.77}$$

$$\beta=\frac{1}{2\pi}\int_{-\infty}^{\ln c}\int_{-\infty}^{\infty}\Big[\prod_{k=1}^{N}\Big(1-\frac{iv}{v_k}\Big)\Big]^{-\frac{1}{2}}\times$$

$$\times\exp\Big\{-\frac{iv}{2}\ln\prod_{k=1}^{N}\Big(1+\frac{1}{v_k}\Big)\Big\}\,e^{ivx}\,dv\,dx, \qquad (5.78)$$

où

$$v_k=\frac{1}{\lambda_k-1}.$$

L'algorithme de détection d'un signal stochastique peut avoir une interprétation analogique dans le cas d'un bruit *blanc* additif, c'est-à-dire pour $B_b(\tau) = N_0\delta(\tau)$. Dans ce cas λ_k sont

$$\lambda_k=1+\frac{1}{N_0\mu_k}, \qquad (5.79)$$

où μ_k sont les nombres caractéristiques de l'équation intégrale linéaire homogène

$$\psi(t)=\mu\int_{-T}^{T}B_s(t-y)\,\psi(y)\,dy, \qquad (5.80)$$

les fonctions propres $\varphi_k(t)$ de l'équation initiale (5.74) étant liées aux fonctions propres $\psi_k(t)$ de l'équation (5.80) par la relation suivante

$$\psi_k(t)=\sqrt{N_0\lambda_k}\,\varphi_k(t). \qquad (5.80')$$

La présence d'un bruit blanc additif exclut la singularité, par conséquent, les sommes dans l'expression (5.75) du rapport de vraisemblance ont des limites finies pour $N\to\infty$ (cf. § 3.5.4). Considérons la somme dans le premier membre de (5.76) pour $N\to\infty$. En substituant à x_k son expression (5.73) et compte tenu de (5.74), (5.80), (5.80') on obtient

$$\sum_{k=1}^{\infty}\frac{\lambda_k-1}{\lambda_k}x_k^2=\frac{1}{N_0}\int_{-T}^{T}\int_{-T}^{T}x(u)\,x(v)\Big[\sum_{k=1}^{\infty}\frac{\psi_k(u)\,\psi_k(v)}{1+N_0\mu_k}\Big]du\,dv. \qquad (5.81)$$

La fonction de deux variables

$$h(u,v)=\sum_{k=1}^{\infty}\frac{\psi_k(u)\,\psi_k(v)}{1+N_0\mu_k} \qquad (5.82)$$

satisfait à l'équation intégrale (cf. § 3.5.6)

$$\int_{-T}^{T}B_s(t-u)\,h(u,v)\,du+N_0h(t,v)=B_s(t-v), \qquad (5.83)$$

$$|t|\leqslant T, \quad |v|\leqslant T.$$

En effet, en portant dans le premier membre de (5.83) l'expression de $h(u, v)$ tirée de (5.82), en changeant l'ordre de la sommation et de l'intégration et compte tenu de (5.80) on obtient

$$\int_{-T}^{T} B_s(t-u)\,h(u, v)\,du = \sum_{k=1}^{\infty} \frac{\psi_k(t)\,\psi_k(v)}{\mu_k(1+N_0\mu_k)} = \sum_{k=1}^{\infty} \frac{\psi_k(t)\,\psi_k(v)}{\mu_k} - N_0 \sum_{k=1}^{\infty} \frac{\psi_k(t)\,\psi_k(v)}{1+N_0\mu_k}.$$

Tenant compte du fait que la première somme est le développement orthogonal de la fonction de corrélation du signal [cf. (3.24)]

$$B_s(t-v) = \sum_{k=1}^{\infty} \frac{\psi_k(t)\,\psi_k(v)}{\mu_k},$$

et que la seconde somme en vertu de (5.82) est égale à $N_0 h(t, v)$ la véridicité de notre affirmation devient évidente.

En utilisant (5.81) et (5.82) on peut formuler le test (5.76) comme suit: le signal est présent dans un bruit blanc additif, si pour la réalisation observée $x(t)$ on a

$$\frac{1}{N_0} \int_{-T}^{T} \int_{-T}^{T} h(u, v)\,x(u)\,x(v)\,du\,dv \geqslant K, \tag{5.84}$$

où

$$K = 2 \ln c + \sum_{i=1}^{\infty} \ln \lambda_i, \tag{5.84'}$$

$h(u, v)$ étant la solution de l'équation intégrale (5.83). Dans (5.84') la somme infinie est convergente et peut s'exprimer en fonction de la *résolvante* du noyau $B_s(t-u)$ de l'équation (5.83) [cf. (3.136)].

Ainsi, le système optimal de détection d'un signal stochastique (normal) noyé dans un bruit blanc normal additif est un filtre non linéaire du second degré (voir § 4.4.2) suivi d'un élément à seuil non inertiel (fig. 5.3) *).

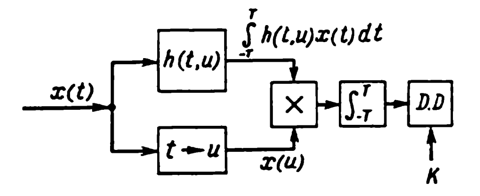


Fig. 5.3. Schéma d'un dispositif optimal de détection d'un signal stochastique

Les probabilités α de fausse alarme et β de perte d'un signal sont, dans le cas considéré, données par les formules (3.141) et (3.141').

*) Comparer également avec le schéma de la figure 3.1.

Considérons en conclusion le cas limite où le signal est également une sorte de « bruit blanc », c'est-à-dire un processus aléatoire à spectre énergétique uniforme pour toutes les fréquences dont l'intensité N_s n'est pas égale à l'intensité N_0 du bruit blanc gênant. On a alors $B_s(\tau) = N_s\delta(\tau)$ et à partir de (5.80) et (5.74) on obtient

$$\mu_k = \frac{1}{N_s}, \quad \lambda_k = 1 + \frac{N_s}{N_0}. \tag{5.85}$$

Le test d'hypothèse d'après N échantillons non corrélés se formule dans ce cas limite comme suit [cf. (5.76)]:
le signal est présent si

$$\frac{N_s/N_0}{1+N_s/N_0} \sum_{k=1}^{N} x_k^2 \geqslant 2 \ln c + N \ln\left(1 + \frac{N_s}{N_0}\right) \tag{5.86}$$

ou

$$\sum_{k=1}^{N} x_k^2 \geqslant \gamma^2, \tag{5.87}$$

avec

$$\gamma^2 = 2\left(1 + \frac{N_0}{N_s}\right) \ln\left[c\left(1 + \frac{N_s}{N_0}\right)^{\frac{N}{2}}\right]. \tag{5.88}$$

Si N est suffisamment grand, la somme dans le premier membre de l'inégalité (5.87) peut être remplacée par une intégrale *)

$$\sum_{k=1}^{N} x_k^2 \sim \frac{1}{N_0} \int_{-T}^{T} x^2(t)\, dt. \tag{5.89}$$

En vertu de (5.87) et (5.89) l'algorithme de détection se réduit à la comparaison du rapport de l'énergie du processus observé $x(t)$ à la densité spectrale du bruit, avec une certaine constante γ^2 déterminée en vertu de (5.88) par la grandeur $\frac{N_s}{N_0}$ et la constante c (c'est-à-dire par le critère choisi). Le dispositif de réception effectuant cette opération peut être appelé *récepteur énergétique*.

Comme x_k sont des variables aléatoires normales indépendantes de moyenne nulle, les variables aléatoires

$$\sum_{k=1}^{N} x_k^2, \quad \frac{N_0}{N_s + N_0} \sum_{k=1}^{N} x_k^2$$

(respectivement en présence et en l'absence de signal) sont réparties suivant la loi du χ_N^2 (cf. problème 3.15, tome I). En utilisant la formule (21) du problème 3.15 on trouve les probabilités condition-

*) Notons que la grandeur γ pour $N \to \infty$ n'est pas limitée.

nelles de fausse alarme et de perte du signal, soit

$$\alpha = 1 - \frac{\Gamma\left(\frac{N}{2}, \frac{\gamma^2}{2}\right)}{\Gamma\left(\frac{N}{2}\right)}, \tag{5.90}$$

$$\beta = \frac{\Gamma\left(\frac{N}{2}, \frac{\gamma^2}{2}\frac{N_0}{N_s + N_0}\right)}{\Gamma\left(\frac{N}{2}\right)}, \tag{5.91}$$

où $\Gamma(x, y)$ est une fonction gamma incomplète [cf. (1.41), tome I].

Si la taille de l'échantillon est importante ($N \gg 1$), en utilisant l'approximation asymptotique de la fonction gamma incomplète [cf. (1.43), tome I] on obtient à partir de (5.90) et (5.91)

$$\alpha \sim 1 - F\left[(\gamma - \sqrt{N})\sqrt{2}\right], \tag{5.92}$$

$$\beta \sim F\left[\left(\gamma\sqrt{\frac{N_0}{N_s + N_0}} - \sqrt{N}\right)\sqrt{2}\right]. \tag{5.93}$$

Les formules (5.92) et (5.93) relient entre eux (pour c donné) le nombre d'observations N, le rapport des puissances du signal N_s et du bruit N_0 par unité de bande de fréquences et les probabilités α de fausse alarme et β de perte du signal.

5.3. DÉTECTION APRÈS DÉMODULATEUR

5.3.1. Méthode d'amplitude. Nous allons passer à l'étude des algorithmes de détection dans le cas ou avant tout traitement le processus à bande étroite reçu est *démodulé*, c'est-à-dire que soit son amplitude, soit sa phase sont extraites du signal modulé. Lors de la réception les processus à haute fréquence sont amplifiés (par exemple, par un amplificateur à fréquence intermédiaire), c'est pourquoi dans ce paragraphe on suppose que la largeur de la bande passante Δ de l'amplificateur prédétecteur soit bien supérieure à la largeur du spectre du signal reçu, à la différence du paragraphe précédent, où le système linéaire était toujours *adapté* au signal.

Considérons tout d'abord le test optimal relatif à la présence ou à l'absence de signal, basé sur la réalisation de l'*enveloppe* du processus observé qui est soit un bruit normal stationnaire à bande étroite de moyenne nulle (hypothèse H_0), soit la somme de ce bruit et d'un signal déterministe à bande étroite $s(t) = a(t)\cos[\omega_0 t - \psi_s(t)]$ (hypothèse H_1). On peut trouver les caractéristiques probabilistes de ces processus dans le chapitre huit du premier tome. En utilisant la terminologie employée dans le chapitre mentionné, on peut formuler comme suit le problème envisagé: il y a lieu de vérifier l'hypothèse simple H_0 que l'enveloppe observée est un processus de

Rayleigh, contre l'alternative simple H_1 que l'enveloppe est un processus de Rayleigh généralisé *).

Selon la méthode générale, il y aurait lieu d'utiliser pour coordonnées observées de l'enveloppe les grandeurs non corrélées

$$r_k = \sqrt{\lambda_k} \int_{-T}^{T} r(t)\, \varphi_k(t)\, dt, \tag{5.94}$$

où $r(t)$ est la réalisation de l'enveloppe sur l'intervalle d'observation $(-T, T)$; λ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale

$$\varphi(t) = \lambda \int_{-T}^{T} B_{\mathrm{E}}(y-t)\, \varphi(y)\, dy, \qquad |t| \leqslant T, \tag{5.94'}$$

$B_{\mathrm{E}}(\tau)$ étant la fonction de corrélation donnée de l'enveloppe du bruit.

Pour trouver la densité de probabilité de la variable aléatoire r_k, il faut, en vertu de (5.94), résoudre l'un des problèmes les plus compliqués de la théorie des processus aléatoires (cf. tome I, chapitre 9) de la détermination de la loi de répartition du processus à la sortie d'un système linéaire, lorsque celle d'entrée n'est pas normale (il s'agit ici de la loi de Rayleigh). Les variables aléatoires r_k, $k = 1, 2, \ldots$, ne sont réparties ni suivant la loi normale, ni suivant la loi de Rayleigh, de sorte que le fait qu'elles ne soient pas corrélées ne signifie pas qu'elles soient indépendantes. Le calcul du rapport de vraisemblance pour l'échantillon $(r_1, \ldots, r_N)$ est un problème fort difficile à résoudre. C'est pourquoi nous allons maintenant nous écarter quelque peu du problème *exact* et faire quelques simplifications **).

Supposant que le spectre énergétique du bruit soit uniforme dans la bande Δ, écrivons les coordonnées non corrélées de l'enveloppe conformément à (3.41), soit:

$$r_k = \sqrt{\frac{\Delta}{\pi N_0}} \int_{-T}^{T} r(t)\, \frac{\sin(\Delta t - \pi k)}{\Delta t - \pi k}\, dt, \tag{5.95}$$

N_0 étant la densité spectrale du bruit.

---

*) Notons que l'enveloppe de la somme d'un processus quasi déterministe (de phase aléatoire) et d'un bruit normal est également répartie suivant une loi de Rayleigh généralisée (cf. § 8.2.1, tome I). C'est pourquoi les conclusions faites dans ce paragraphe concernent également la détection d'un signal quasi déterministe [cf. (5.30)].

**) Seule cette méthode simpliste (sans fondements rigoureux) était utilisée dans les premiers travaux de la théorie de la détection des signaux (voir, par exemple, [18]).

Supposons de plus que le temps d'observation $T \gg \frac{1}{\Delta}$. Comme pour $\Delta \to \infty$ la fonction $\frac{\Delta}{\pi} \frac{\sin(\Delta t - \pi k)}{\Delta t - \pi k} \to \delta\left(t - \frac{\pi k}{\Delta}\right)$, compte tenu de (5.95) et vu que $\Delta T \gg 1$, on a

$$r_k = \sqrt{\frac{\pi}{N_0 \Delta}}\, r\left(\frac{\pi k}{\Delta}\right) \tag{5.96}$$

ou

$$r_k = \frac{E_k}{\sigma}, \tag{5.96'}$$

$\sigma^2 = \frac{N_0 \Delta}{\pi}$ étant la variance du bruit; $E_k = r\left(\frac{\pi k}{\Delta}\right)$ les valeurs échantillonnées de l'enveloppe pour $t = \frac{\pi k}{\Delta}$.

Ainsi, l'échantillonnage de l'enveloppe est fait à des intervalles de temps $\frac{\pi}{\Delta}$, les valeurs échantillonnées pouvant *approximativement* être supposées non corrélées.

Comme dans le cas d'un processus aléatoire normal les valeurs non corrélées de l'enveloppe sont statistiquement indépendantes (cf. tome I, § 8.2), les coordonnées non corrélées r_k sont des variables aléatoires *indépendantes.*

En se limitant aux N premières coordonnées, écrivons les fonctions de vraisemblance de l'échantillon $(r_1, \ldots, r_N)$ pour les deux hypothèses mentionnées ci-dessus [cf. (8.17) et (8.18), tome I]:

$$W_N(r_1, \ldots, r_N \mid H_0) = \prod_{k=1}^{N} r_k e^{-\frac{r_k^2}{2}}, \quad r_k > 0, \tag{5.97}$$

$$W_N(r_1, \ldots, r_N \mid H_1) = \prod_{k=1}^{N} r_k e^{-\frac{r_k^2 + a_k^2}{2}} I_0(r_k a_k), \quad r_k > 0, \tag{5.97'}$$

où

$$a_k = \frac{a\left(\frac{\pi k}{\Delta}\right)}{\sigma} \tag{5.98}$$

est le rapport de l'amplitude du signal, à l'instant $t = \frac{\pi k}{\Delta}$, à la moyenne quadratique du bruit.

Trouvons à partir de (5.97) et (5.97') le logarithme du rapport de vraisemblance, soit:

$$\ln l(r_1, \ldots, r_N) = \sum_{k=1}^{N} \ln I_0(r_k a_k) - \frac{1}{2} \sum_{k=1}^{N} a_k^2. \tag{5.99}$$

Maintenant nous pouvons facilement formuler le test d'hypothèse optimal: si pour l'échantillon observé $r_1, \ldots, r_N$ des coordonnées

de l'enveloppe on a

$$\sum_{k=1}^{N} \ln I_0(r_k a_k) \geqslant \ln c + \frac{1}{2} \sum_{k=1}^{N} a_k^2 = K_N, \tag{5.100}$$

on adopte la décision γ_1 que le signal est présent, et si c'est l'inégalité inverse de (5.100) qui est vérifiée, on adopte la décision γ_0 qu'il n'y a pas de signal.

Dans la formule (5.100) la constante c est déterminée par le critère de qualité adopté (de Bayes, de Neumann-Pearson, du minimax) conformément à la table 1 page 35.

On ne peut dans ce cas calculer sous une forme générale les probabilités conditionnelles des fausses alarmes et des pertes du signal. Cependant on peut étudier l'expression approchée de ces probabilités pour un signal faible lorsque $\max_k a_k \ll 1$.

En développant $\ln I_0(r_k a_k)$ en série suivant les puissances de $(r_k a_k)$ et en nous limitant au premier terme, c'est-à-dire en négligeant les termes contenant les puissances supérieures ou égales à la quatrième *) nous avons:

$$\ln I_0(r_k a_k) \approx \frac{1}{4}(r_k a_k)^2.$$

On obtient alors à partir de (5.100) le test simplifié : le signal est présent si

$$\sum_{k=1}^{N} a_k^2 r_k^2 \geqslant 4K_N. \tag{5.101}$$

Ainsi, l'algorithme de détection se réduit dans le cas présent au calcul de la somme pondérée des carrés des valeurs échantillonnées de l'enveloppe et à sa comparaison avec un seuil dépendant seulement du critère choisi et des caractéristiques a priori du signal et du bruit.

En utilisant (3.6) et (3.7) et leur généralisation éventuelle aux processus aléatoires (cf. § 3.2.3), on peut approximativement, pour $T\Delta \gg 1$, remplacer la somme (5.101) par une intégrale

$$\sum_{k=1}^{N} a_k^2 = \frac{1}{\sigma^2} \sum_{k=1}^{N} a^2\left(\frac{\pi k}{\Delta}\right) \approx \frac{E_a}{N_0}, \tag{5.102}$$

où

$$E_a = \int_{-T}^{T} a^2(t)\,dt \tag{5.103}$$

et

$$\frac{1}{4} \sum_{k=1}^{N} r_k^2 a_k^2 = \frac{1}{4\sigma^4} \sum_{k=1}^{N} r^2\left(\frac{\pi k}{\Delta}\right) a^2\left(\frac{\pi k}{\Delta}\right) \approx \frac{\Delta}{4\pi\sigma^4} \int_{-T}^{T} r^2(t)\, a^2(t)\,dt. \tag{5.104}$$

*) Le développement en série et l'approximation faite doivent recevoir une interprétation probabiliste.

En portant (5.102) et (5.104) dans (5.101) on obtient pour le test en question l'expression analytique suivante:

$$\frac{1}{4N_0\sigma^2}\int_{-T}^{T} r^2(t)\,a^2(t)\,dt \gtrless \ln c + \frac{E_a}{2N_0}, \tag{5.105}$$

qui montre que les éléments essentiels d'un dispositif optimal de réception pour la méthode d'amplitude de détection d'un signal faible sont un détecteur quadratique et un corrélomètre calculant la fonction de corrélation mutuelle du signal à la sortie du détecteur $r^2(t)$ et du carré de l'enveloppe $a^2(t)$ du signal déterministe.

Pour trouver les probabilités de fausse alarme et de perte du signal en se basant sur (5.101) ou (5.105), on a dans le cas général à résoudre un problème assez compliqué de la répartition de la somme pondérée des carrés des variables aléatoires indépendantes, réparties suivant la loi de Rayleigh généralisée dans le premier cas, et de la répartition de l'intégrale du carré du processus de Rayleigh généralisé, dans le second.

Si l'échantillon est de taille assez grande ($N \gg 1$), la répartition de la somme de variables aléatoires indépendantes dans le premier membre de (5.100) est, en vertu du théorème limite central et pour les restrictions faites concernant a_k (cf. § 3.4, tome I), normale, de moyenne et de variance égales respectivement à *) (pour $\max\limits_k a_k \ll 1$)

$$m_1\left\{\sum_{k=1}^{N}\ln I_0(a_k r_k)\,|\,H_0\right\} \approx 2\sum_{k=1}^{N} a_k^2\left(1-\frac{a_k^2}{2}\right), \tag{5.106}$$

$$M_2\left\{\sum_{k=1}^{N}\ln I_0(a_k r_k)\,|\,H_0\right\} \approx 4\sum_{k=1}^{N} a_k^4, \tag{5.106'}$$

en l'absence de signal et

$$m_1\left\{\sum_{k=1}^{N}\ln I_0(a_k r_k)\,|\,H_1\right\} \approx 2\sum_{k=1}^{N} a_k^2\left(1+\frac{a_k^2}{2}\right), \tag{5.106''}$$

$$M_2\left\{\sum_{k=1}^{N}\ln I_0(a_k r_k)\,|\,H_1\right\} \approx 4\sum_{k=1}^{N} a_k^4, \tag{5.106'''}$$

en présence du signal.

En vertu de (5.106)-(5.106'''), pour $\max\limits_k a_k \ll 1$ et $N \gg 1$ le logarithme du rapport de vraisemblance (5.99) est réparti (approxi-

*) Pour le calcul des moments des carrés de variables aléatoires réparties suivant la loi de Rayleigh et la loi de Rayleigh généralisée voir tome I, chapitres 2 et 3.

mativement) suivant une loi normale de paramètres

$$m_1\{\ln l(r_1, \ldots, r_N) \mid H_0\} = -\frac{1}{4}\sum_{k=1}^{N} a_k^4, \tag{5.107}$$

$$m_1\{\ln l(r_1, \ldots, r_N) \mid H_1\} = \frac{1}{4}\sum_{k=1}^{N} a_k^4, \tag{5.107'}$$

$$M_2\{\ln l(r_1, \ldots, r_N) \mid H_0\} = M_2\{\ln l(r_1, \ldots, r_N) \mid H_1\} = \frac{1}{4}\sum_{k=1}^{N} a_k^4. \tag{5.107''}$$

En introduisant la notation

$$s_N^2 = \frac{1}{4}\sum_{k=1}^{N} a_k^4, \tag{5.108}$$

on peut écrire comme suit les probabilités de fausse alarme et de perte du signal:

$$\alpha = 1 - F\left(\frac{\ln c}{s_N} + s_N\right), \tag{5.109}$$

$$\beta = F\left(\frac{\ln c}{s_N} - s_N\right). \tag{5.109'}$$

Notons que pour la méthode absolument optimale de détection, les probabilités de fausse alarme et de perte du signal [cf. (5.14) et (5.15)] dépendent du paramètre d_N égal à la racine carrée de la somme des carrés des valeurs normées du signal, alors que les probabilités données par les formules (5.109) et (5.109') dépendent du paramètre s_N égal à la racine carrée de la somme des quatrièmes puissances des valeurs normées de l'enveloppe du signal. Ceci veut dire que, pour un signal faible, lors d'un traitement optimal après démodulation il y a altération de la caractéristique de la détection par rapport au traitement optimal avant démodulation, en effet, dans le premier cas c'est le carré du rapport signal/bruit qui entre en jeu, alors que dans le second c'est ce rapport lui-même.

On peut étudier d'une manière analogue les propriétés asymptotiques de la caractéristique de la détection à partir de l'algorithme (5.105) de traitement de la réalisation pour $T \to \infty$ et compte tenu du fait que la répartition de l'intégrale dans (5.105) tend vers la loi normale (cf. § 9.4.2, tome I).

Notons que pour un signal *intense*, c'est-à-dire pour $\min_k a_k \gg 1$, en utilisant le développement asymptotique de la fonction de Bessel (cf., par exemple, § 3.2.2, tome I) on peut écrire comme suit l'inégalité (5.100)

$$\sum_{k=1}^{N} a_k r_k \gg K_N \tag{5.110}$$

et pour $T\Delta \gg 1$ [cf. (5.105)] on a approximativement

$$\frac{1}{N_0}\int_{-T}^{T} r(t)\,a(t)\,dt \geqslant \ln c + \frac{E_a}{2N_0}. \tag{5.111}$$

Dans ce cas le dispositif de réception optimal se compose d'un détecteur linéaire et d'un corrélomètre donnant la fonction de corrélation mutuelle entre le signal à la sortie du détecteur et l'enveloppe du signal déterministe.

Comme précédemment, pour $N \gg 1$, on trouve que le logarithme du rapport de vraisemblance pour $\min_k a_k \gg 1$ suit une loi normale de paramètres

$$m_1\{\ln l(r_1, \ldots, r_N) \mid H_0\} \sim -\frac{1}{2}\sum_{k=1}^{N} a_k^2, \tag{5.112}$$

$$m_1\{\ln l(r_1, \ldots, r_N) \mid H_1\} \sim \frac{1}{2}\sum_{k=1}^{N} a_k^2, \tag{5.112'}$$

$$M_2\{\ln l(r_1, \ldots, r_N) \mid H_0\} \sim 2\sum_{k=1}^{N} a_k^2, \tag{5.112''}$$

$$M_2\{\ln l(r_1, \ldots, r_N) \mid H_1\} \sim \sum_{k=1}^{N} a_k^2. \tag{5.112'''}$$

Les probabilités de fausse alarme et de perte du signal sont respectivement

$$\alpha = 1 - F\left(\frac{\ln c}{d_N\sqrt{2}} + \frac{d_N}{2\sqrt{2}}\right), \tag{5.113}$$

$$\beta = F\left(\frac{\ln c}{d_N} - \frac{d_N}{2}\right), \tag{5.114}$$

où

$$d_N^2 = \sum_{k=1}^{N} a_k^2. \tag{5.115}$$

Notons que dans le cas d'un signal intense ($d_N \gg 1$) la probabilité de perte du signal donnée par (5.114) est la même que dans le cas du traitement avant démodulation [cf. (5.15)]. Quant à la probabilité de fausse alarme, le second terme (le terme essentiel) dans l'argument de l'intégrale de Laplace est $\sqrt{2}$ fois moindre que le terme correspondant dans la formule (5.14).

5.3.2. **Méthode de phase.** Cherchons maintenant le test optimal permettant de juger de la présence ou l'absence d'un signal déterministe noyé dans un bruit normal additif à bande étroite d'après la réalisation de la *phase* $\vartheta(t)$ du processus observé. Les caractéristiques probabilistes de la phase correspondant aux deux hypothèses

H_0 (phase du bruit) et H_1 (phase du mélange additif du signal et du bruit) sont données dans le chapitre huit du premier tome.

En répétant des raisonnements analogues à ceux ayant conduit à la formule (5.96) introduisons les coordonnées non corrélées de la phase

$$\vartheta_k = \vartheta\left(\frac{\pi k}{\Delta}\right). \tag{5.116}$$

Les valeurs de la phase étant non corrélées et appartenant à un processus normal, les coordonnées ϑ_k sont indépendantes (cf. § 8.4.2, tome I). En se limitant aux N premières coordonnées on peut écrire comme suit les fonctions de vraisemblance de l'échantillon $\vartheta_1, \ldots$ $\ldots, \vartheta_N$ pour les deux hypothèses mentionnées [cf. (8.59), tome I]:

$$w_N(\vartheta_1, \ldots, \vartheta_N \mid H_0) = \left(\frac{1}{2\pi}\right)^N, \quad |\vartheta_i| \leqslant \pi, \tag{5.117}$$

$$w_N(\vartheta_1, \ldots, \vartheta_N \mid H_1) = \prod_{k=1}^{N} \left\{\frac{1}{2\pi} e^{-\frac{a_k^2}{2}} + \right.$$
$$+ \frac{a_k}{\sqrt{2\pi}} \cos(\vartheta_k - \psi_k) F[a_k \cos(\vartheta_k - \psi_k)] \times$$
$$\left. \times e^{-\frac{a_k^2}{2}\sin^2(\vartheta_k - \psi_k)}\right\}, \quad |\vartheta_i - \psi_i| \leqslant \pi, \quad i = 1, \ldots, N, \tag{5.118}$$

où a_k est donné par la formule (5.98) et

$$\psi_k = \psi_s\left(\frac{\pi k}{\Delta}\right). \tag{5.119}$$

A partir de (5.117) et (5.118) on peut obtenir le logarithme du rapport de vraisemblance, soit:

$$\ln(\vartheta_1, \ldots, \vartheta_N) = \sum_{k=1}^{N} \ln \{e^{-\frac{a_k^2}{2}} + \sqrt{2\pi}\, a_k \cos(\vartheta_k - \psi_k) \times$$
$$\times F[a_k \cos(\vartheta_k - \psi_k)]\, e^{-\frac{a_k^2}{2}\sin^2(\vartheta_k - \psi_k)}\}, \tag{5.120}$$
$$|\vartheta_i - \psi_i| \leqslant \pi, \quad i = 1, \ldots, N.$$

Le test d'hypothèse peut alors se formuler comme suit: le signal est présent si

$$\sum_{k=1}^{N} \ln \{e^{-\frac{a_k^2}{2}} + \sqrt{2\pi}\, a_k \cos(\vartheta_k - \psi_k) \times$$
$$\times F[a_k \cos(\vartheta_k - \psi_k)]\, e^{-\frac{a_k^2}{2}\sin^2(\vartheta_k - \psi_k)}\} \geqslant \ln c, \tag{5.121}$$
$$|\vartheta_i - \psi_i| \leqslant \pi, \quad i = 1, \ldots, N.$$

Dans le cas d'un signal *faible* ($\max_k a_k \ll 1$) la relation (5.121) se simplifie considérablement. En utilisant la formule (8.61) du premier tome on trouve en première approximation

$$\sqrt{\frac{\pi}{2}} \sum_{k=1}^{N} a_k \cos(\vartheta_k - \psi_k) \geqslant \ln c + \frac{\pi}{8} \sum_{k=1}^{N} a_k^2 = K_N, \tag{5.122}$$

$$|\vartheta_i - \psi_i| \leqslant \pi, \qquad i = 1, \ldots, N.$$

Comme dans les cas précédents, pour $\Delta T \gg 1$ la somme dans le premier membre de (5.122) peut approximativement être remplacée par une intégrale

$$\frac{\Delta}{\pi\sigma} \int_{-T}^{T} a(t) \cos[\vartheta(t) - \psi_s(t)]\, dt \geqslant \sqrt{\frac{2}{\pi}} \ln c + \sqrt{\frac{\pi}{32}} \frac{E_a}{2N_0}. \tag{5.122'}$$

Ainsi, l'opération principale de l'algorithme optimal de détection d'un signal déterministe faible par la méthode de phase est le calcul de la fonction de corrélation mutuelle entre le cosinus de la différence de phase du signal reçu et du signal déterministe et l'enveloppe du signal déterministe *).

Tout comme dans la méthode d'amplitude, le calcul des probabilités de fausse alarme et de perte du signal d'après l'algorithme de traitement (5.122) se heurte à des difficultés du calcul de la densité de probabilité de la somme des variables aléatoires du premier membre de (5.122). Cependant, si l'échantillon est de taille importante ($N \gg 1$) on peut supposer que cette somme est asymptotiquement normale et se limiter au calcul de la moyenne et de la variance pour les deux hypothèses. Pour $\max_k a_k \ll 1$ (cf. § 8.5.1, tome I) on a

$$m_1 \left\{ \sqrt{\frac{\pi}{2}} \sum_{k=1}^{N} a_k \cos(\vartheta_k - \psi_k) \mid H_0 \right\} = 0, \tag{5.123}$$

$$m_1 \left\{ \sqrt{\frac{\pi}{2}} \sum_{k=1}^{N} a_k \cos(\vartheta_k - \psi_k) \mid H_1 \right\} = \frac{\pi}{4} \sum_{k=1}^{N} a_k^2, \tag{5.123'}$$

$$M_2 \left\{ \sqrt{\frac{\pi}{2}} \sum_{k=1}^{N} a_k \cos(\vartheta_k - \psi_k) \mid H_0 \right\} =$$

$$= M_2 \left\{ \sqrt{\frac{\pi}{2}} \sum_{k=1}^{N} a_k \cos(\vartheta_k - \psi_k) \mid H_1 \right\} = \frac{\pi}{4} \sum_{k=1}^{N} a_k^2. \tag{5.123''}$$

*) Comme il a été montré dans [23], le dispositif optimal de détection par la méthode de phase est un limiteur parfait suivi d'un filtre adapté.

En utilisant l'expression (5.115), on trouve pour les probabilités de fausse alarme et de perte du signal :

$$\alpha = 1 - F\left(\frac{2\ln c}{\sqrt{\pi}\,d_N} + \frac{\sqrt{4}}{4}\,d_N\right), \tag{5.124}$$

$$\beta = F\left(\frac{2\ln c}{\sqrt{\pi}\,d_N} - \frac{\sqrt{\pi}}{4}\,d_N\right). \tag{5.124'}$$

A la différence de la méthode optimale d'amplitude [cf. (5.109) et (5.109')], dans la méthode optimale de phase, tout comme dans la méthode optimale avant démodulation, les probabilités de fausse alarme et de perte du signal dépendent du paramètre d_N, c'est-à-dire que la caractéristique de détection d'un signal faible se trouve être meilleure dans la méthode de phase que dans la méthode d'amplitude *). On peut facilement trouver l'analogie existant entre (5.124), (5.124') et (5.14), (5.15), les premières formules pouvant être obtenues à partir des secondes en substituant $\sqrt{\frac{\pi}{4}}\,d_N$ (c'est-à-dire $\approx$ $\approx 0{,}9\,d_N$) à d_N.

A l'aide de la formule asymptotique (8.63) du premier tome, dans le cas d'un signal *intense* ($\min\limits_k a_k \gg 1$), le test d'hypothèse [cf. (5.121)] peut se formuler comme suit : le signal est présent si

$$\frac{1}{2}\sum_{k=1}^{N} a_k^2(\vartheta_k - \psi_k)^2 \leqslant \sum_{k=1}^{N} \ln\left(\sqrt{2\pi}\,a_k c^{-\frac{1}{N}}\right) \tag{5.125}$$

ou approximativement, pour $\Delta T \gg 1$, si

$$\frac{1}{2N_0}\int_{-T}^{T} a^2(t)\,[\vartheta(t) - \psi_s(t)]^2\,dt \leqslant \frac{\Delta}{\pi}\int_{-T}^{T} \ln\left[\sqrt{2\pi}\,\frac{a(t)}{\sigma}\right]dt - \ln c. \tag{5.125'}$$

Ainsi, le dispositif optimal de réception, pour la méthode de détection de phase d'un signal déterministe intense discriminant se compose essentiellement d'un détecteur de phase, séparant le carré de la différence de phase des signaux reçu et déterministe, et d'un corrélomètre donnant la fonction de corrélation mutuelle du signal à la sortie du détecteur de phase et le carré de l'enveloppe du signal.

Considérons également la méthode optimale de détection de phase appliquée à un signal quasi déterministe [cf. (5.30)] de phase initiale aléatoire. Dans ce cas, pour un signal faible ($\max\limits_k a_k \ll 1$), le rapport de vraisemblance pour une valeur donnée de la phase φ_0 est

*) On pouvait s'y attendre car la densité de probabilité de la phase (5.118) est plus riche en information que la densité de probabilité de l enveloppe (5.97'). Dans le premier cas les paramètres de la répartition recèlent l'information sur les valeurs de l'enveloppe et de la phase du signal, tandis que dans le second. seulement sur les valeurs de l'enveloppe.

[cf. (5.122)]

$$l(\vartheta_1, \ldots, \vartheta_N \mid \varphi_0) =$$
$$= \exp\left\{\sqrt{\frac{\pi}{2}}\sum_{k=1}^{N} a_i \cos(\vartheta_k - \psi_k + \varphi_0) - \frac{\pi}{8}\sum_{k=1}^{N} a_k^2\right\} \quad (5.126)$$

ou

$$l(\vartheta_1, \ldots, \vartheta_N \mid \varphi_0) = \exp\left\{\sqrt{\frac{\pi}{2}}\, r_N \cos(\varphi_0 - \Phi_N)\right\} \times$$
$$\times \exp\left\{-\frac{\pi}{8}\sum_{k=1}^{N} a_k^2\right\}, \quad (5.126')$$

où

$$r_N = \left|\sum_{k=1}^{N} a_k e^{i(\vartheta_k - \psi_k)}\right|; \quad (5.127)$$

$$\Phi_N = \operatorname{arc\,tg}\frac{\operatorname{Im}\sum_{k=1}^{N} a_k e^{i(\vartheta_k - \psi_k)}}{\operatorname{Re}\sum_{k=1}^{N} a_k e^{i(\vartheta_k - \psi_k)}}. \quad (5.127')$$

Prenons la moyenne du rapport de vraisemblance par rapport à la phase aléatoire φ_0 uniformément répartie, il vient:

$$\Lambda(\vartheta_1, \ldots, \vartheta_N) = \exp\left\{-\frac{\pi}{8}\sum_{k=1}^{N} a_k^2\right\} \times$$
$$\times \frac{1}{2\pi}\int_0^{2\pi} \exp\left\{\sqrt{\frac{\pi}{2}}\, r_N \cos(\varphi_0 - \Phi_N)\right\} d\varphi_0 =$$
$$= I_0\left(\sqrt{\frac{\pi}{2}}\, r_N\right) \exp\left\{-\frac{\pi}{8}\sum_{k=1}^{N} a_k^2\right\}. \quad (5.128)$$

De la formule (5.128) découle le test de Bayes suivant: le signal quasi déterministe est présent dans le processus observé si

$$\ln I_0\left(\sqrt{\frac{\pi}{2}}\, r_N\right) \geqslant \ln c + \frac{\pi}{8}\sum_{k=1}^{N} a_k^2 = K_N \quad (5.128')$$

ou bien, si

$$r_N \geqslant c_N$$

c_N étant donnée par l'équation transcendante

$$\ln I_0\left(\sqrt{\frac{\pi}{2}}\, c_N\right) = K_N. \quad (5.128'')$$

On peut calculer sous forme explicite la densité de probabilité de la statistique r_N [cf. (5.127)] pour l'hypothèse H_0, c'est-à-dire en l'absence de signal. Dans ce cas, en utilisant avec quelques modifications la méthode examinée dans [5], chapitre 4, § 4, relativement au problème du « voyageur errant », on obtient l'expression suivante pour la densité de probabilité unidimensionnelle de r_N, lorsque ϑ_k sont des variables aléatoires indépendantes uniformément réparties sur l'intervalle $(0, 2\pi)$:

$$w_1(r) = r \int_0^\infty sJ_0(rs) \prod_{k=1}^{N} J_0(a_k s)\, ds.$$

Il en découle, compte tenu de (5.128′), que la probabilité de fausse alarme est

$$\alpha = \int_{c_N}^\infty w_1(r)\, dr = 1 - c_N \int_0^\infty J_1(c_N s) \prod_{k=1}^{N} J_0(a_k s)\, ds. \tag{5.129}$$

Pour un signal sinusoïdal d'amplitude constante on a

$$\alpha = 1 - k \int_0^\infty J_1(kx)\, J_0^N(x)\, dx, \tag{5.129′}$$

où $k = \frac{c_N}{A_0}$. Pour le critère de Neumann-Pearson le seuil k est donné par (5.129′) en fonction de la probabilité α de fausse alarme et ne dépend pas de la puissance du signal.

Pour un échantillon de taille importante ($N \gg 1$) la loi de répartition de la variable aléatoire r_N tend asymptotiquement vers la loi de Rayleigh pour l'hypothèse H_0 et vers la loi de Rayleigh généralisée pour l'hypothèse H_1 (cf. [5] pages 187-188), de paramètres respectifs [cf. (5.123)-(5.123″)] $\frac{1}{2}\, d_N^2$ et $\left(\sqrt{\frac{\pi}{8}}\, d_N^2,\ \frac{1}{2}\, d_N^2\right)$, d_N^2 étant donné par la formule (5.115). Les probabilités de fausse alarme et de perte du signal sont alors [cf. (5.52) et (5.53)]

$$\alpha = e^{-\frac{c_N^2}{d_N^2}}. \tag{5.130}$$

$$\beta = e^{-\frac{\pi}{8} d_N^2} \int_0^{\frac{c_N \sqrt{2}}{d_N}} x e^{-\frac{x^2}{x}} I_0\left(\sqrt{\frac{\pi}{4}}\, x d_N\right) dx. \tag{5.130′}$$

5.3.3. Signal stochastique. Considérons maintenant la méthode de détection optimale d'amplitude d'un signal stochastique en conservant les hypothèses faites au début du paragraphe 5.3. Les fonc-

tions de vraisemblance seront dans ce cas [cf. (5.97)]

$$W_N(r_1, \ldots, r_N \mid H_0) = \prod_{k=1}^{N} r_k e^{-r_k^2/2}, \qquad r_k \geqslant 0,$$

$$W_N(r_1, \ldots, r_N \mid H_1) = \prod_{k=1}^{N} \frac{r_k}{1+\frac{\sigma_s^2}{\sigma^2}} e^{-\frac{r_k^2}{2\left(1+\frac{\sigma_s^2}{\sigma^2}\right)}},$$

$$r_k \geqslant 0, \ k=1, \ldots, N,$$

σ_s^2 étant la variance du signal.

Le rapport de vraisemblance est

$$l(r_1, \ldots, r_N) = \left(1+\frac{\sigma_s^2}{\sigma^2}\right)^{-N} \exp\left\{\frac{\sigma_s^2/\sigma^2}{2(1+\sigma_s^2/\sigma^2)} \sum_{k=1}^{N} r_k^2\right\},$$

$$r_k \geqslant 0, \quad k=1, \ldots, N. \tag{5.131}$$

Le test d'hypothèse se formule comme suit : le signal est présent si la somme des carrés des valeurs échantillonnées r_k est non inférieure à un certain seuil [cf. (5.86) et (5.87)]

$$\sum_{k=1}^{N} r_k^2 \geqslant \gamma^2, \tag{5.132}$$

où γ est la même constante que dans (5.87) calculée à l'aide de la formule (5.88) en remplaçant évidemment $\frac{N_s}{N_0}$ par $\left(\frac{\sigma_s}{\sigma}\right)^2$.

Comme r_k^2 sont des variables aléatoires indépendantes réparties suivant une loi exponentielle, en utilisant le résultat du problème 3.16 du premier tome, on trouve que $\sum_{k=1}^{N} r_k^2$ suit une loi du χ_{2N}^2. En tenant compte du fait que l'échantillon des valeurs de l'enveloppe du processus est deux fois plus petit que l'échantillon du processus lui-même (cf. § 5.1), on voit que les formules (5.90) et (5.91) pour les probabilités de fausse alarme et de perte du signal *ne changent pas dans le cas de la méthode de détection optimale d'amplitude.* Ainsi, les procédures optimales de détection d'un signal stochastique d'après les méthodes avant et après démodulation coïncident. On pouvait s'attendre à ce résultat. En effet, dans le cas envisagé les fonctions de vraisemblance des échantillons de phase en présence et en l'absence de signal sont *les mêmes*

$$W_N(\vartheta_1, \ldots, \vartheta_N \mid H_0) = W_N(\vartheta_1, \ldots, \vartheta_N \mid H_1) = \left(\frac{1}{2\pi}\right)^N,$$

du fait que la densité de probabilité unidimensionnelle de la phase d'un processus aléatoire normal stationnaire ne dépend pas du spectre

énergétique (de la variance) du processus, étant toujours égale à $\frac{1}{2\pi}$, $|\vartheta_i| \leqslant \pi$. Le rapport de vraisemblance est dans ce cas identiquement égal à l'unité. Par conséquent, dans le cas envisagé on ne peut pas, d'après les échantillons indépendants de phase, distinguer un bruit pur de la somme d'un signal et d'un bruit, c'est-à-dire résoudre le problème de la détection d'un signal stochastique.

Pour réaliser la méthode de détection de phase d'un signal stochastique il est indispensable d'avoir des échantillons indépendants de différence de phases *) $\Delta\vartheta_1, \ldots, \Delta\vartheta_N$ où $\Delta\vartheta_i = \vartheta(t_i + \tau) - \vartheta(t_i)$, $\tau > 0$. Utilisons la formule (8.82) du premier tome pour trouver la fonction de vraisemblance d'un échantillon de différence de phases pour un bruit

$$W_N(\Delta\vartheta_1, \ldots, \Delta\vartheta_N \mid H_0) = \left(\frac{1-R_0^2}{2\pi}\right)^N \times$$
$$\times \prod_{k=1}^{N} \left[\frac{1}{1-y_k^2} + \frac{\frac{\pi}{2} + \arcsin y_k}{(1-y_k^2)^{3/2}}\, y_k\right], \tag{5.133}$$
$$|\Delta\vartheta_k| \leqslant \pi, \quad k = 1, \ldots, N,$$

où

$$y_k = R_0(\tau)\cos[\vartheta(t_k+\tau) - \vartheta(t_k)]; \tag{5.134}$$
$$R_0^2(\tau) = R_{c0}^2(\tau) + R_{s0}^2(\tau); \tag{5.134'}$$
$$R_{c0}(\tau) = \frac{1}{\sigma^2}\int_0^\infty F_b(\omega)\cos(\omega-\omega_0)\tau\, d\omega; \tag{5.135}$$
$$R_{s0}(\tau) = \frac{1}{\sigma^2}\int_0^\infty F_b(\omega)\sin(\omega-\omega_0)\tau\, d\omega; \tag{5.135'}$$

$F_b(\omega)$, σ^2 étant le spectre énergétique et la variance du bruit.

La fonction de vraisemblance de la somme du signal et du bruit est

$$W_N(\Delta\vartheta_1, \ldots, \Delta\vartheta_N \mid H_1) = \left(\frac{1-R_1^2}{2\pi}\right)^N \times$$
$$\times \prod_{k=1}^{N} \left[\frac{1}{1-z_k^2} + \frac{\frac{\pi}{2} + \arcsin z_k}{(1-z_k^2)^{3/2}}\, r_k\right], \tag{5.136}$$
$$|\Delta\vartheta_k| \leqslant \pi, \quad k = 1, \ldots, N,$$

où

$$z_k = R_1(\tau)\cos[\vartheta(t_k+\tau) - \vartheta(t_k)]; \tag{5.137}$$
$$R_1^2(\tau) = R_{c1}^2(\tau) + R_{s1}^2(\tau); \tag{5.137'}$$

*) Ceci a lieu, par exemple, dans les problèmes de la réception en diversité.

$$R_{c1}(\tau)=\frac{1}{\sigma_s^2+\sigma^2}\int_0^\infty [F_s(\omega)+F_b(\omega)]\cos(\omega-\omega_0)\tau\,d\omega; \quad (5.138)$$

$$R_{s1}(\tau)=\frac{1}{\sigma_s^2+\sigma^2}\int_0^\infty [F_s(\omega)+F_b(\omega)]\sin(\omega-\omega_0)\tau\,d\omega; \quad (5.138')$$

$F_s(\omega)$, σ_s^2 étant respectivement le spectre énergétique et la variance du signal.

Le rapport de vraisemblance est

$$l(\Delta\vartheta_1,\ldots,\Delta\vartheta_N)=\left(\frac{1-R_1^2}{1-R_0^2}\right)^N\times$$

$$\times\prod_{k=1}^N\left[\frac{1-y_k^2}{1-z_k^2}\,\frac{1+z_k\left(\frac{\pi}{2}+\arcsin z_k\right)(1-z_k^2)^{-\frac{1}{2}}}{1+y_k\left(\frac{\pi}{2}+\arcsin y_k\right)(1-y_k^2)^{-\frac{1}{2}}}\right], \quad (5.139)$$

$$|\Delta\vartheta_i|\leqslant\pi,\quad i=1,\ldots,N.$$

Le cas général est très difficile à étudier, c'est pourquoi nous allons nous limiter au cas où les échantillons de bruit $\vartheta(t_k+\tau)$ et $\vartheta(t_k)$ ne sont pas corrélés, c'est-à-dire où $R_0(\tau)\equiv 0$, $\tau>0$ (cf. (8.81), tome I) et $\sigma_s^2\ll\sigma^2$ (signal *faible*). Introduisons la notation

$$R^2(\tau)=R_c^2(\tau)+R_s^2(\tau), \quad (5.140)$$

où

$$R_c(\tau)=\frac{1}{\sigma_s^2}\int_0^\infty F_s(\omega)\cos(\omega-\omega_0)\tau\,d\omega, \quad (5.141)$$

$$R_s(\tau)=\frac{1}{\sigma_s^2}\int_0^\infty F_s(\omega)\sin(\omega-\omega_0)\tau\,d\omega, \quad (5.141')$$

et utilisons le fait que le rapport $\left(\frac{\sigma_s}{\sigma}\right)^2$ est petit, il vient

$$R_1(\tau)=\frac{\sigma_s^2}{\sigma^2}R(\tau). \quad (5.142)$$

Posons dans (5.139) $R_0\equiv 0$ et tenons compte de (5.142), il vient

$$\ln l(\Delta\vartheta_1,\ldots,\Delta\vartheta_N)\approx\frac{\pi R}{2}\left(\frac{\sigma_s}{\sigma}\right)^2\sum_{k=1}^N\cos\Delta\vartheta_k-$$

$$-\frac{N\pi^2R^2}{16}\left(\frac{\sigma_s}{\sigma}\right)^4,\quad |\vartheta_i|\leqslant\pi,\quad i=1,\ldots,N. \quad (5.143)$$

Le test d'hypothèse se formule alors comme suit : le signal est présent si

$$\sum_{k=1}^{N} \cos \Delta\vartheta_k \geqslant \frac{2}{\pi R} \left(\frac{\sigma}{\sigma_s}\right)^2 \ln c + \frac{N\pi R}{8} \left(\frac{\sigma_s}{\sigma}\right)^2. \qquad (5.144)$$

La procédure optimale de détection consiste à comparer le signal à la sortie de l'intégrateur (additionneur) des cosinus des différences de phases avec un seuil donné à l'avance.

Si l'échantillon est de taille importante ($N \gg 1$) la loi de répartition de la somme des variables aléatoires indépendantes $\cos \Delta\vartheta_k$ est voisine d'une loi normale de moyenne et de variance égales à (cf. § 8.4.4, tome I)

$$m_1 \left\{\sum_{k=1}^{N} \cos \Delta\vartheta_k \mid H_0\right\} = 0, \qquad (5.145)$$

$$m_1 \left\{\sum_{k=1}^{N} \cos \Delta\vartheta_k \mid H_1\right\} = \frac{N\pi R}{4} \left(\frac{\sigma_s}{\sigma}\right)^2, \qquad (5.145')$$

$$M_2 \left\{\sum_{k=1}^{N} \cos \Delta\vartheta_k \mid H_0\right\} = M_2 \left\{\sum_{k=1}^{N} \cos \Delta\vartheta_k \mid H_1\right\} = \frac{N}{2}. \qquad (5.146)$$

Les probabilités de fausse alarme et de perte du signal ont pour expressions *)

$$\alpha = 1 - F\left[\frac{4 \ln c}{\pi R \sqrt{2N}} \left(\frac{\sigma}{\sigma_s}\right)^2 + \frac{\pi R}{8} \left(\frac{\sigma_s}{\sigma}\right)^2 \sqrt{2N}\right], \qquad (5.147)$$

$$\beta = F\left[\frac{4 \ln c}{\pi R \sqrt{2N}} \left(\frac{\sigma}{\sigma_s}\right)^2 - \frac{\pi R}{8} \left(\frac{\sigma_s}{\sigma}\right)^2 \sqrt{2N}\right]. \qquad (5.148)$$

5.4. DISCRIMINATION DES SIGNAUX

5.4.1. Notes préliminaires. Jusqu'à présent nous avons envisagé seulement le cas de la détection d'un signal d'après une certaine réalisation du processus à l'entrée d'un récepteur pouvant être soit un bruit, soit la somme d'un signal et d'un bruit. Dans de nombreuses applications on voit apparaître des problèmes plus compliqués. Par exemple, le processus observé $x(t)$ peut être la somme d'un bruit et de l'un des signaux appartenant à un ensemble donné de signaux $s_0(t), \ldots, s_m(t)$. Le problème se pose comme suit : d'après une

*) Comparer (5.147) et (5.148) avec (6) et (6') dans le problème (5.5). Si l'on tient compte du fait que dans le cas envisagé la taille totale de l'échantillon est $2N$, on voit que la caractéristique de détection d'un signal stochastique, pour les restrictions mentionnées ci-dessus, tend vers la caractéristique de détection d'un signal déterministe par la méthode d'amplitude, de paramètre $\frac{\pi R}{4}$ fois plus petit $\left(\text{pour } a = \frac{\sigma_s}{\sigma}\right)$.

réalisation du processus $x(t)$ observé sur l'intervalle $(-T, T)$ décider lequel des signaux mentionnés est réellement présent dans le processus observé.

Un problème analogue de discrimination des signaux apparaît en télécommunications, lorsqu'il s'agit des dispositifs de décodage des communications. Dans d'autres cas il faut dire si le processus observé est uniquement un bruit, ou si c'est un mélange additif du bruit avec un ou plusieurs signaux de l'ensemble donné de signaux. Dans la radiodétection (radar) ces problèmes sont parfois appelés problèmes de *distinction des signaux*. Dans le cas général les problèmes mentionnés se rapportent à la classe des problèmes à alternatives multiples (cf. § 1.4.9). Nous allons étudier dans ce qui suit quelques problèmes relativement simples de ce type, leur solution étant cependant intéressante par elle-même pour certaines applications pratiques.

5.4.2. Deux signaux déterministes *). Supposons que l'on sache que le processus observé est un mélange additif d'un bruit normal stationnaire et de l'un des deux signaux déterministes $s_0(t)$ ou $s_1(t)$. Il s'agit de vérifier l'hypothèse simple H_0 que le processus observé est réparti suivant une loi normale de moyenne $s_0(t)$, contre l'alternative simple H_1 que ce processus est également réparti suivant une loi normale mais de moyenne $s_1(t)$.

Conformément au § 3.4, choisissons en tant que coordonnées observées les coordonnées non corrélées

$$x_k = \sqrt{\lambda_k} \int_{-T}^{T} x(t)\, \varphi_k(t)\, dt, \tag{5.149}$$

$x(t)$ étant une réalisation du processus sur l'intervalle d'observation $(-T, T)$; λ_k et $\varphi_k(t)$, les nombres caractéristiques et les fonctions propres de l'équation intégrale

$$\varphi(t) = \lambda \int_{-T}^{T} B(y-t)\, \varphi(y)\, dy, \tag{5.150}$$

où $B(\tau)$ est la fonction de corrélation donnée du bruit.

En se limitant aux N premières coordonnées et compte tenu de (3.62), on peut écrire comme suit le logarithme du rapport de vraisemblance pour l'échantillon $x_1, \ldots, x_N$ de taille N

$$\ln l(x_1, \ldots, x_N) = \frac{1}{2} \sum_{k=1}^{N} (a_k^2 - b_k^2) + \sum_{k=1}^{N} (b_k - a_k)\, x_k, \tag{5.151}$$

*) Ce paragraphe peut être considéré comme une généralisation du § 5.2.1, dont les résultats sont un cas particulier pour $s_0(t) \equiv 0$ et $s_1(t) \equiv s(t)$.

où

$$a_k = \sqrt{\lambda_k} \int_{-T}^{T} s_0(t)\, \varphi_k(t)\, dt\,; \tag{5.152}$$

$$b_k = \sqrt{\lambda_k} \int_{-T}^{T} s_1(t)\, \varphi_k(t)\, dt. \tag{5.152'}$$

On aura alors le test d'hypothèse suivant: on décide de la présence du signal $s_1(t)$ si pour l'échantillon observé on a

$$\sum_{k=1}^{N} (b_k - a_k)\, x_k \geqslant \ln c + \frac{1}{2} \sum_{k=1}^{N} (b_k^2 - a_k^2), \tag{5.153}$$

et de la présence du signal $s_0(t)$ si c'est l'inégalité inverse de (5.153) qui est vérifiée.

Ainsi, l'algorithme de discrimination se réduit au calcul de la somme pondérée des coordonnées observées et à sa comparaison avec un seuil établi à l'avance:

$$K_N = \ln c + \frac{1}{2} \sum_{k=1}^{N} (b_k^2 - a_k^2). \tag{5.154}$$

Cet algorithme est optimal par rapport à tous les critères de qualité (sauf les critères d'échantillonnage progressif) mentionnés dans le premier chapitre. Le choix du critère de qualité n'influe que sur la grandeur c.

En répétant les raisonnements du § 5.2.1 on obtient les probabilités conditionnelles d'erreurs de première et de seconde espèce (probabilités de *confusion* des signaux)

$$\alpha = 1 - F\left(\frac{\ln c}{d_N} + \frac{1}{2} d_N\right), \tag{5.155}$$

$$\beta = F\left(\frac{\ln c}{d_N} - \frac{1}{2} d_N\right), \tag{5.156}$$

où

$$d_N^2 = \sum_{k=1}^{N} (b_k - a_k)^2. \tag{5.157}$$

Si les résultats des observations ne peuvent être échantillonnés et si toute la réalisation $x(t)$ se trouve utilisée sur l'intervalle d'observation, le test d'hypothèse se formule comme suit [cf. (3.81)]: dans le processus le signal $s_1(t)$ est présent si

$$\int_{-T}^{T} V(t)\, x(t)\, dt \geqslant \ln c + \frac{1}{2} \int_{-T}^{T} V(t)\, [s_0(t) + s_1(t)]\, dt = K_T, \tag{5.158}$$

et le signal $s_0(t)$ pour l'inégalité inverse de (5.158). La fonction $V(t)$ est la solution de l'équation intégrale linéaire non homogène

$$\int_{-T}^{T} B(t-u)\,V(u)\,du = s_1(t) - s_0(t). \qquad (5.159)$$

Les probabilités de confusion des signaux sont données par les formules (5.155) et (5.156) où l'on remplace d_N par d_T donné par la relation [cf. (3.87) et (3.88)]

$$d_T^2 = \int_{-T}^{T} V(t)\,[s_1(t) - s_0(t)]\,dt. \qquad (5.160)$$

Lors de la discrimination des signaux noyés dans un bruit *blanc* additif la solution de l'équation intégrale (5.159) est

$$V(t) = \frac{1}{N_0}\,[s_1(t) - s_0(t)], \qquad (5.161)$$

où N_0 est la densité spectrale du bruit. Dans ce cas l'inégalité (5.158) peut s'écrire comme suit :

$$\int_{-T}^{T} [s_1(t) - s_0(t)]\,x(t)\,dt \geqslant N_0 \ln c + \\ + \frac{1}{2}\int_{-T}^{T} [s_1^2(t) - s_0^2(t)]\,dt = N_0 \ln c + \frac{E_1 - E_0}{2}, \qquad (5.162)$$

où E_1 et E_0 sont les énergies des signaux $s_1(t)$ et $s_0(t)$ sur l'intervalle d'observation.

Après transformations algébriques simples l'inégalité (5.162) devient

$$\int_{-T}^{T} [s_0(t) - x(t)]^2\,dt - \int_{-T}^{T} [s_1(t) - x(t)]^2\,dt \geqslant N_0 \ln c. \qquad (5.162')$$

Les intégrales dans (5.162') peuvent être appelées *distances* respectives entre la réalisation observée $x(t)$ et les signaux $s_0(t)$ et $s_1(t)$. Le test d'hypothèse indique qu'il faut calculer la différence de ces distances et la comparer avec un seuil dépendant du critère de qualité et de la densité spectrale du bruit blanc. Pour le critère du maximum de vraisemblance ($c = 1$) on adopte la décision que c'est le signal *le plus proche* de la réalisation observée qui est présent. Cette règle est à la base du récepteur parfait de Kotelnikov. La caractéristique de fonctionnement de ce récepteur est donnée par la grandeur [cf. (5.160)]

$$d_T^2 = \frac{1}{N_0}\int_{-T}^{T} [s_1(t) - s_0(t)]^2\,dt, \qquad (5.163)$$

c'est-à-dire par le rapport de la distance entre les signaux à la densité spectrale du bruit. Ayant trouvé à l'aide de (5.163) la grandeur d_T^2, on peut utiliser pour les calculs les courbes données sur la figure 5.1 [cf. (5.155) et (5.156)].

Notons que pour des signaux *orthogonaux* pour lesquels on a

$$\int_{-T}^{T} s_1(t)\, s_0(t)\, dt \equiv 0,$$

la grandeur d_T^2 est égale à $\frac{E_1+E_2}{N_0}$, c'est-à-dire au rapport de l'énergie totale des signaux à la densité spectrale du bruit $\left(\text{pour } E_1=E_2=E \text{ on a } d_T^2 = \frac{2E}{N_0}\right)$. La valeur maximale possible de d_T^2 est égale à $\frac{4E}{N_0}$ et correspond à $s_1(t) = -s_0(t)$.

En vertu de (5.162), le récepteur optimal pour la discrimination de deux signaux déterministes noyés dans un bruit blanc normal additif se compose de *deux filtres adaptés* (ou de deux corrélomètres, cf. § 4.3.4) de fonctions de transition impulsionnelles

$$h_0^*(\tau) = s_0(T-\tau), \quad |\tau| \leqslant T, \tag{5.164}$$

$$h_1^*(\tau) = s_1(T-\tau), \quad |\tau| \leqslant T, \tag{5.164'}$$

suivis de dispositifs de sélection, de soustraction et de coupure. Pour des signaux de même énergie ($E_1 = E_2$), lorsque l'on utilise le critère du maximum de vraisemblance, les deux derniers dispositifs sont remplacés par un dispositif de comparaison, car dans ce cas le test (5.162) devient: le signal $s_1(t)$ est présent si

$$\int_{-T}^{T} s_1(t)\, x(t)\, dt \geqslant \int_{-T}^{T} s_0(t)\, x(t)\, dt. \tag{5.165}$$

5.4.3. Deux signaux d'amplitudes inconnues. Supposons maintenant que l'on connaisse les signaux à un facteur d'amplitude A près. Le problème de la discrimination des signaux est maintenant dans la vérification de l'hypothèse *composite* H_0 selon laquelle le processus observé suit une loi normale de moyenne $As_0(t)$ contre l'alternative *composite* H_1 que la moyenne de ce processus normal est $As_1(t)$.

Supposons que nous n'ayons aucune information a priori sur le paramètre A. En utilisant les résultats du § 3.5.3 il est facile de voir que le logarithme de la fonctionnelle du rapport de vraisemblance peut s'écrire comme suit:

$$\ln l\,[x(t)] = A \int_{-T}^{T} V(t)\, x(t)\, dt - \frac{A^2}{2} \int_{-T}^{T} V(t)\,[s_0(t) + s_1(t)]\, dt, \tag{5.166}$$

où $V(t)$ est la solution de l'équation intégrale (5.159).

Par conséquent, pour $A > 0$, pour le test uniformément le plus puissant l'hypothèse H_0 est rejetée, c'est-à-dire que l'on affirme que c'est le signal $As_1(t)$ qui est présent, si pour la réalisation observée $x(t)$ on a

$$\int_{-T}^{T} V(t)\, x(t)\, dt \geqslant K_T. \tag{5.167}$$

Le seuil K_T est donné par la probabilité α d'erreur de première espèce

$$K_T = d_T x_\alpha, \tag{5.167'}$$

où x_α est le quantile de la loi normale, d_T étant donné par (5.160).

Si $A < 0$ le sens de l'inégalité (5.167) se trouve changé, et la grandeur K_T remplacée par $-K_T$.

En vertu du test uniformément le plus puissant non biaisé il y a lieu de rejeter l'hypothèse H_0 si

$$\left| \int_{-T}^{T} V(t)\, x(t)\, dt \right| \geqslant K_T. \tag{5.168}$$

Ainsi, le dispositif optimal (d'après le critère de Neumann-Pearson) de discrimination de deux signaux d'amplitudes inconnues noyés dans un bruit blanc additif est en principe le même que dans le cas de signaux donnés.

En particulier, pour $s_0(t) \equiv 0$, en vertu de (5.167) l'algorithme optimal de détection du signal déterministe $As(t)$ d'amplitude inconnue A se réduit à la comparaison de la grandeur $\int_{-T}^{T} V(t)\, x(t)\, dt$ avec un seuil, $V(t)$ étant la solution de l'équation (5.22) [coïncidant avec (5.159) pour $s_1(t) \equiv s(t)$ et $s_0(t) \equiv 0$]. C'est ce qu'impose le test (5.21) lors de la détection du signal déterministe $s(t)$. Les caractéristiques probabilistes de la détection dépendront évidemment de la grandeur A, car dans ce cas le paramètre d_T [cf. (5.23)] sera

$$d_T^2 = A^2 \int_{-T}^{T} V(t)\, s(t)\, dt.$$

Indiquons un autre critère qui n'est pas lié à l'information a priori sur le paramètre inconnu, c'est le critère du maximum de vraisemblance (cf. § 1.4.3). En vertu de (1.116), dans ce cas il faut d'abord trouver les maxima de la fonction de vraisemblance pour les deux hypothèses et comparer leur rapport avec l'unité. En prenant pour coordonnées observées les grandeurs x_k données par (5.149) et en conservant les notations (5.152), (5.152'), on peut écrire comme suit

les fonctions de vraisemblance de l'échantillon $x_1, \ldots, x_N$ pour les deux hypothèses concurrentes *):

$$W_N(x_1, \ldots, x_N \mid H_0) = \frac{1}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2}\sum_{k=1}^{N}(x_k - Aa_k)^2\right], \quad (5.169)$$

$$W_N(x_1, \ldots, x_N \mid H_1) = \frac{1}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2}\sum_{k=1}^{N}(x_k - Ab_k)^2\right]. \quad (5.169')$$

Les fonctions $f(A)$ et $\ln f(A)$ ayant des maxima pour une même valeur de A, il est facile de montrer que le maximum de la fonction (5.169) a lieu pour

$$\hat{A}_0 = \frac{\sum_{k=1}^{N} a_k x_k}{\sum_{k=1}^{N} a_k^2}, \quad (5.170)$$

et celui de la fonction (5.169′) pour

$$\hat{A}_1 = \frac{\sum_{k=1}^{N} b_k x_k}{\sum_{k=1}^{N} b_k^2}. \quad (5.170')$$

En remplaçant A dans (5.169) par la grandeur $\hat{A}_0$, et dans (5.169′) par la grandeur $\hat{A}_1$, on obtient

$$\ln \frac{\max_A W_N(x_1, \ldots, x_N \mid H_1)}{\max_A W_N(x_1, \ldots, x_N \mid H_0)} = \frac{1}{2d_{N1}^2}\left(\sum_{k=1}^{N} b_k x_k\right)^2 - \frac{1}{2d_{N0}^2}\left(\sum_{k=1}^{N} a_k x_k\right)^2, \quad (5.171)$$

où

$$d_{N0}^2 = \sum_{k=1}^{N} a_k^2; \quad (5.172)$$

$$d_{N1}^2 = \sum_{k=1}^{N} b_k^2. \quad (5.172')$$

Le test d'hypothèse basé sur le critère du maximum de vraisemblance peut se formuler comme suit: c'est le signal $As_1(t)$ qui est présent si

$$\frac{1}{d_{N1}^2}\left(\sum_{k=1}^{N} b_k x_k\right)^2 \geqslant \frac{1}{d_{N0}^2}\left(\sum_{k=1}^{N} a_k x_k\right)^2, \quad (5.173)$$

et le signal $As_0(t)$ pour l'inégalité inverse de (5.173).

*) Rappelons qu'au § 5.2.1 nous avons montré que $M_2\{x_i\} = 1$.

Lorsque les énergies des signaux sont égales ($d_{N0} = d_{N1}$) la relation (5.173) devient (5.153) pour $c = 1$.

En utilisant les résultats du § 3.4.4, il est facile d'écrire la relation (5.173) sous forme analogique, soit:

$$\frac{1}{d_{T1}^2}\Big(\int_{-T}^{T} V_1(t)\,x(t)\,dt\Big)^2 \geqslant \frac{1}{d_{T0}^2}\Big(\int_{-T}^{T} V_0(t)\,x(t)\,dt\Big)^2, \qquad (5.174)$$

où

$$d_{T1}^2 = \int_{-T}^{T} V_1(t)\,s_1(t)\,dt\,; \qquad (5.175)$$

$$d_{T0}^{\cdot\cdot} = \int_{-T}^{T} V_0(t)\,s_0(t)\,dt, \qquad (5.175')$$

et $V_1(t)$, $V_0(t)$ sont les solutions des équations intégrales

$$\int_{-T}^{T} B(t-u)\,V_1(u)\,du = s_1(t), \quad |t| \leqslant T, \qquad (5.176)$$

$$\int_{-T}^{T} B(t-u)\,V_0(t)\,du = s_0(t), \quad |t| \leqslant T. \qquad (5.176')$$

Le récepteur optimal se compose dans ce cas de deux filtres adaptés de fonctions de transition impulsionnelles

$$h_0^*(\tau) = \frac{1}{d_{T0}}\,V_0(T-\tau), \quad |\tau| \leqslant T, \qquad (5.177)$$

$$h_1^*(\tau) = \frac{1}{d_{T1}}\,V_1(T-\tau), \quad |\tau| \leqslant T, \qquad (5.177')$$

suivis de dispositifs de sélection, d'élévation au carré et de comparaison.

Dans le cas de la discrimination de deux signaux noyés dans un bruit blanc les solutions des équations intégrales (5.177), (5.177') sont

$$V_0(t) = \frac{1}{N_0}\,s_0(t), \quad V_1(t) = \frac{1}{N_0}\,s_1(t),$$

et (5.174) donne le test suivant: c'est le signal $As_1(t)$ qui est présent si

$$\frac{1}{E_1}\Big(\int_{-T}^{T} s_1(t)\,x(t)\,dt\Big)^2 \geqslant \frac{1}{E_0}\Big(\int_{-T}^{T} s_0(t)\,x(t)\,dt\Big)^2, \qquad (5.178)$$

E_0 et E_1 étant les énergies des signaux $s_0(t)$ et $s_1(t)$ sur l'intervalle d'observation. Pour $E_0 = E_1$ l'inégalité (5.178) devient évidemment (5.165).

Si l'on se donne la densité de probabilité a priori $w_1(A)$ du paramètre A, on peut utiliser le critère de Bayes de risque moyen minimal. Dans ce cas conformément au § 1.4.2 on calcule le rapport des fonctions de vraisemblance centrées [cf. (5.169) et (5.169')]

$$\Lambda(x_1, \ldots, x_N) = \frac{\int\limits_{-\infty}^{\infty} w_1(A) \exp\left[-\frac{1}{2}\sum\limits_{k=1}^{N} (x_k - Ab_k)^2\right] dA}{\int\limits_{-\infty}^{\infty} w_1(A) \exp\left[-\frac{1}{2}\sum\limits_{k=1}^{N} (x_k - Aa_k)^2\right] dA} \tag{5.179}$$

et suivant que cette grandeur surpasse ou non le seuil c (donné par les probabilités a priori d'apparition des signaux et de coût des décisions) on adopte la décision sur la présence de l'un des signaux $As_1(t)$ ou $As_0(t)$.

A titre d'exemple, nous allons considérer le cas où la densité de probabilité a priori de l'amplitude A est normale de moyenne A_0 et de variance σ_A^2, c'est-à-dire

$$w_1(A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{(A-A_0)^2}{2\sigma_A^2}\right]. \tag{5.180}$$

Portons (5.180) dans (5.179) et calculons les intégrales figurant dans le numérateur et le dénominateur, il vient

$$\ln\Lambda(x_1, \ldots, x_N) = \frac{1}{2}\ln\frac{1+\sigma_A^2 d_{N1}^2}{1+\sigma_A^2 d_{N0}^2} + \frac{A_0^2}{2} \times \\ \times \left(\frac{d_{N0}^2}{1+\sigma_A^2 d_{N0}^2} - \frac{d_{N1}^2}{1+\sigma_A^2 d_{N1}^2}\right) + A_0 \sum_{k=1}^{N}\left(\frac{b_k}{1+\sigma_A^2 d_{N1}^2} - \right. \\ \left. - \frac{a_k}{1+\sigma_A^2 d_{N0}^2}\right) x_k + \frac{\sigma_A^2}{2}\left[\frac{1}{1+\sigma_A^2 d_{N1}^2}\left(\sum_{k=1}^{N} b_k x_k\right)^2 - \right. \\ \left. - \frac{1}{1+\sigma_A^2 d_{N0}^2}\left(\sum_{k=1}^{N} a_k x_k\right)^2\right], \tag{5.181}$$

où d_{N0}^2 et d_{N1}^2 sont donnés par (5.172) et (5.172').

Le test de Bayes se formule alors comme suit : c'est le signal $As_1(t)$ qui est présent si

$$A_0 \sum_{k=1}^{N}\left(\frac{b_k}{1+\sigma_A^2 d_{N1}^2} - \frac{a_k}{1+\sigma_A^2 d_{N0}^2}\right) x_k + \frac{\sigma_A^2}{2}\left[\frac{1}{1+\sigma_A^2 d_{N1}^2}\left(\sum_{k=1}^{N} b_k x_k\right)^2 - \right. \\ \left. - \frac{1}{1+\sigma_A^2 d_{N0}^2}\left(\sum_{k=1}^{N} a_k x_k\right)^2\right] \geqslant K_N, \tag{5.182}$$

où

$$K_N = \ln c + \frac{A_0^2}{2}\left(\frac{d_{N1}^2}{1+\sigma_A^2 d_{N1}^2} - \frac{d_{N0}^2}{1+\sigma_A^2 d_{N0}^2}\right) + \frac{1}{2}\ln\frac{1+\sigma_A^2 d_{N0}^2}{1+\sigma_A^2 d_{N1}^2}. \qquad (5.182')$$

Pour $\sigma_A \to 0$ la fonction $w_1(A) \to \delta(A - A_0)$ et l'inégalité (5.182) devient

$$A_0 \sum_{k=1}^{N} (b_k - a_k)\, x_k \geqslant \ln c + \frac{A_0^2}{2}(d_{N1}^2 - d_{N0}^2), \qquad (5.183)$$

par conséquent, ce test coïncide avec le test de Bayes de discrimination de signaux entièrement connus (avec la condition évidente $A_0 = 1$).

Pour $\sigma_A \to \infty$ on trouve à partir de (5.182)

$$\frac{1}{2d_{N1}^2}\left(\sum_{k=1}^{N} b_k x_k\right)^2 - \frac{1}{2d_{N0}^2}\left(\sum_{k=1}^{N} a_k x_k\right)^2 \geqslant \ln c + \frac{1}{2}\ln\frac{d_{N0}^2}{d_{N1}^2}. \qquad (5.184)$$

Les opérations essentielles auxquelles doivent être soumises les coordonnées obvervées, sont ici les mêmes que dans le cas du critère du maximum de vraisemblance [cf. (5.173)]. La différence est en ce qu'au lieu d'une simple comparaison entre elles des sorties des dispositifs d'élévation au carré, leurs différences sont comparées avec un certain seuil fixé à l'avance.

Pour des signaux de même énergie ($d_{N0} = d_{N1}$) les formules (5.182) à (5.184) se simplifient comme suit

$$A_0 \sum_{k=1}^{N} (b_k - a_k)\, x_k + \frac{\sigma_A^2}{2} \sum_{k=1}^{N} (b_k^2 - a_k^2)\, x_k^2 \geqslant (1+\sigma_A^2)\ln c,$$

$$A_0 \sum_{k=1}^{N} (b_k - a_k)\, x_k \geqslant \ln c \quad \text{pour } \sigma_A \to 0,$$

$$\sum_{k=1}^{N} (b_k^2 - a_k^2)\, x_k^2 \geqslant 2\ln c \quad \text{pour } \sigma_A \to \infty.$$

Notons que le processus quasi déterministe $As(t)$ d'amplitude aléatoire A est souvent utilisé dans les applications comme modèle d'un signal noyé dans un bruit multiplicatif (stationnaire), *variant lentement* durant la période d'observation (ou sur un intervalle égal à la durée du signal impulsionnel).

5.4.4. Deux signaux à bande étroite de phases aléatoires *). Nous allons maintenant étudier le problème de la discrimination de deux signaux quasi déterministes à bande étroite noyés dans un bruit

*) Les résultats du § 5.2.2 sont un cas particulier pour $s_0(t) \equiv 0$, $s_1(t) \equiv s(t)$ des résultats figurant ici.

blanc additif dont les amplitudes et les phases sont modulées suivant une loi donnée et dont les phases initiales sont aléatoires, indépendantes et uniformément réparties, soit:

$$s_0(t) = a_0(t) \cos[\omega_0 t - \psi_{s0}(t) + \varphi_0], \qquad (5.185)$$

$$s_1(t) = a_1(t) \cos[\omega_1 t - \psi_{s1}(t) + \varphi_1]. \qquad (5.185')$$

Supposons que l'on ait

$$0 < \Delta = \frac{\omega_1 - \omega_0}{2} \ll \omega_0.$$

Exprimons les processus à bande étroite en fonction de l'enveloppe et de la phase, ayant choisi pour pulsation centrale

$$\omega^* = \frac{\omega_0 + \omega_1}{2}.$$

En répétant les raisonnements du § 5.2.2, nous pouvons dans le cas envisagé obtenir l'expression suivante pour le logarithme du rapport des fonctions moyennes de vraisemblance de N coordonnées non corrélées de l'enveloppe complexe [cf. (5.42)]:

$$\Lambda(z_1, \ldots, z_N) = \exp\left\{-\frac{1}{2}\sum_{k=1}^{N}[|s_{k1}|^2 - |s_{k0}|^2]\right\} \times \frac{\int_0^{2\pi} \exp[r_{N1}\cos(\varphi_1 - \psi_{N1})]\,d\varphi_1}{\int_0^{2\pi} \exp[r_{N0}\cos(\varphi_0 - \psi_{N0})]\,d\varphi_0}, \qquad (5.186)$$

où

$$r_{N1} = \left|\sum_{k=1}^{N} z_k \overline{s_{k1}}\right|; \quad \psi_{N1} = \operatorname{arc\,tg}\frac{\operatorname{Im}\sum_{k=1}^{N} z_k \overline{s_{k1}}}{\operatorname{Re}\sum_{k=1}^{N} z_k s_{k1}}; \qquad (5.187)$$

$$r_{N0} = \left|\sum_{k=1}^{N} z_k \overline{s_{k0}}\right|; \quad \psi_{N0} = \operatorname{arc\,tg}\frac{\operatorname{Im}\sum_{k=1}^{N} z_k \overline{s_{k0}}}{\operatorname{Re}\sum_{k=1}^{N} z_k \overline{s_{k0}}}. \qquad (5.187')$$

Les grandeurs s_{k0} et s_{k1} sont les coordonnées des enveloppes complexes des signaux [cf. (5.39‴)] données par les formules

$$s_{k0} = \sqrt{\lambda_k}\int_{-T}^{T} z_{s0}(t)\, e^{-it\Delta}\overline{\varphi_k(t)}\, dt, \qquad (5.188)$$

$$s_{k1} = \sqrt{\lambda_k}\int_{-T}^{T} z_{s1}(t)\, e^{it\Delta}\overline{\varphi_k(t)}\, dt, \qquad (5.188')$$

où

$$z_{s0}(t) = a_0(t)\, e^{-i\psi_{s0}(t)}; \quad z_{s1}(t) = a_1(t)\, e^{-i\psi_{s1}(t)}.$$

Intégrons (5.186) et prenons le logarithme, il vient

$$\ln \Lambda(z_1, \ldots, z_N) = \ln I_0(r_{N1}) - \ln I_0(r_{N0}) - \frac{1}{2} \sum_{k=1}^{N} [|s_{k1}|^2 - |s_{k0}|^2], \tag{5.189}$$

où $I_0(r)$ est la fonction de Bessel d'ordre zéro de l'argument imaginaire.

Le test de Bayes basé sur la relation (5.189) se formule comme suit : on adopte la décision γ_1 que c'est le signal $s_1(t)$ qui est présent si pour l'échantillon observé on a l'inégalité

$$\ln I_0(r_{N1}) - \ln I_0(r_{N0}) \geqslant \ln c + \frac{1}{2} \sum_{k=1}^{N} [|s_{k1}|^2 - |s_{k0}|^2] = K_N, \tag{5.190}$$

et la déciison γ_0 affirmant la présence du signal $s_0(t)$ pour l'inégalité inverse.

Si la discrimination s'effectue non pas d'après un échantillon discret, mais d'après la réalisation $z(t)$ de l'enveloppe complexe observée sur l'intervalle $(-T, T)$, en vertu de (5.190) pour $N \to \infty$ [cf. (5.62)] on a

$$\ln I_0(r_{T1}) - \ln I_0(r_{T0}) \geqslant \ln c + \frac{d_{T1}^2 - d_{T0}^2}{2} = K_T, \tag{5.191}$$

où

$$r_{T1} = \left| \int_{-T}^{T} U_1(t)\, \overline{z(t)}\, dt \right|, \tag{5.192}$$

$$d_{T1}^2 = \int_{-T}^{T} U_1(t)\, \overline{z_{s1}(t)}\, e^{-it\Delta}\, dt, \tag{5.192'}$$

$$r_{T0} = \left| \int_{-T}^{T} U_0(t)\, \overline{z(t)}\, dt \right|, \tag{5.193}$$

$$d_{T0}^2 = \left| \int_{-T}^{T} U_0(z)\, \overline{z_{s0}(t)}\, e^{it\Delta}\, dt \right|, \tag{5.193'}$$

U_1 (t) et U_0 (t) étant les solutions des équations intégrales linéaires non homogènes

$$\int_{-T}^{T} B_z(t-u)\, U_1(u)\, du = z_{s1}(t)\, e^{it\Delta}, \quad |t| \leqslant T, \tag{5.194}$$

$$\int_{-T}^{T} B_z(t-u)\, U_0(u)\, du = z_{s0}(t)\, e^{-it\Delta}, \quad |t| \leqslant T, \tag{5.194'}$$

où z_{s1} (t) et z_{s0} (t) sont les enveloppes complexes des signaux s_1 (t) et s_0 (t), et B_z (τ), tout comme dans (5.35′), la fonction de corrélation de l'enveloppe complexe du bruit.

Les variables aléatoires r_{T1}, r_{T0} dans l'inégalité (5.191) sont réparties suivant la loi de Rayleigh généralisée, car ce sont les modules des vecteurs plans dont les composantes sont indépendantes et normalement réparties. Les moyennes et les variances de ces vecteurs (qui sont des variables aléatoires complexes), compte tenu de (5.194) et (5.194′), sont

$$m_{10} = m_1 \left\{ \int_{-T}^{T} U_1(t)\, \overline{z(t)}\, dt \mid H_0 \right\} = \int_{-T}^{T}\int_{-T}^{T} B_z(t-u)\, U_1(t)\, \overline{U_0(u)}\, dt\, du, \tag{5.195}$$

$$m_{11} = m_1 \left\{ \int_{-T}^{T} U_1(t)\, \overline{z(t)}\, dt \mid H_1 \right\} = \int_{-T}^{T}\int_{-T}^{T} B_z(t-u)\, U_1(t)\, \overline{U_1(u)}\, dt\, du, \tag{5.195'}$$

$$m_{00} = m_1 \left\{ \int_{-T}^{T} U_0(t)\, \overline{z(t)}\, dt \mid H_0 \right\} = \int_{-T}^{T}\int_{-T}^{T} B_z(t-u)\, U_0(t)\, \overline{U_0(u)}\, dt\, du, \tag{5.195''}$$

$$m_{01} = m_1 \left\{ \int_{-T}^{T} U_0(t)\, \overline{z(t)}\, dt \mid H_1 \right\} = \int_{-T}^{T}\int_{-T}^{T} B_z(t-u)\, U_0(t)\, \overline{U_1(u)}\, dt\, du, \tag{5.196}$$

$$M_{21} = M_2 \left\{ \int_{-T}^{T} U_1(t)\, \overline{z(t)}\, dt \mid H_0, H_1 \right\} =$$

$$= \int_{-T}^{T}\int_{-T}^{T} B_z(t-u)\, U_1(t)\, \overline{U_1(t)}\, dt\, du, \tag{5.196'}$$

$$M_{20} = M_2 \left\{ \int_{-T}^{T} U_0(t)\, \overline{z(t)}\, dt \mid H_0, H_1 \right\} =$$

$$= \int_{-T}^{T}\int_{-T}^{T} B_z(t-u)\, U_0(t)\, \overline{U_0(u)}\, dt\, du. \tag{5.196''}$$

A titre d'exemple nous allons étudier le cas simple de la discrimination de deux signaux noyés dans un bruit *blanc* additif reçus en manipulation de fréquence lorsque $a_1(t) = a_0(t) = A_0$, $\psi_{s1}(t) = \psi_{s0}(t) = 0$. Nous allons utiliser le critère du maximum de vraisemblance ($c = 1$). Le test (5.191) se trouve maintenant simplifié : on décide que c'est le signal $s_1(t) = A_0 \cos(\omega_1 t + \varphi_1)$ qui est présent si

$$\ln I_0(r_{T1}) \geqslant \ln I_0(r_{T0}), \tag{5.197}$$

et le signal $s_0(t) = A_0 \cos(\omega_0 t + \varphi_0)$ pour l'inégalité inverse. Comme le logarithme de la fonction de Bessel modifiée $\ln I_0(r)$ est une fonction monotone de l'argument r, le test (5.197) revient à établir laquelle des grandeurs r_{T1} ou r_{T0} est plus grande. Ainsi, nous adoptons que c'est le signal $s_1(t)$ qui est présent si

$$r_{T1} \geqslant r_{T0}. \tag{5.197'}$$

Ecrivons sous forme explicite les grandeurs figurant dans (5.197'), tenant compte du fait que dans le cas envisagé on a

$$z_{s1}(t) = z_{s0}(t) \equiv A_0$$

et que pour un bruit blanc de densité spectrale N_0 les solutions des équations (5.194) et (5.194') sont

$$U_0(t) = \frac{A_0}{N_0} e^{-it\Delta}, \quad U_1(t) = \frac{A_0}{N_0} e^{it\Delta}.$$

On obtient, à partir des relations (5.197), (5.192) et (5.193) le test suivant de discrimination des signaux : on adopte que c'est le signal de pulsation ω_1 qui est présent si on a

$$\left| \int_{-T}^{T} \overline{z(t)}\, e^{i\frac{\omega_1-\omega_0}{2}}\, dt \right| \geqslant \left| \int_{-T}^{T} \overline{z(t)}\, e^{-i\frac{\omega_1-\omega_0}{2}} \right|. \tag{5.198}$$

En exprimant l'enveloppe complexe en fonction des composantes en quadrature [cf. (5.32')] on peut écrire (5.198) comme suit

$$\int_{-T}^{T} A(t) \cos\frac{\omega_1-\omega_0}{2} t\, dt \int_{-T}^{T} C(t) \sin\frac{\omega_1-\omega_0}{2} t\, dt \geqslant$$
$$\geqslant \int_{-T}^{T} A(t) \sin\frac{\omega_1-\omega_0}{2} t\, dt \int_{-T}^{T} C(t) \cos\frac{\omega_1-\omega_0}{2} t\, dt. \tag{5.198'}$$

Comme $A(t)$ et $C(t)$ sont des fonctions variant lentement (cf. § 6.2.2, tome I), négligeant les intégrales du produit de ces fonctions par les porteuses haute fréquence du type $\cos \omega t$ (pour $\omega \geqslant \omega_0$) le test (5.198') peut s'exprimer en fonction de la réalisation du processus reçu

$$x(t) = A(t) \cos\frac{\omega_0+\omega_1}{2} t + C(t) \sin\frac{\omega_0+\omega_1}{2} t.$$

On adopte la décision que c'est le signal de pulsation ω_1 qui est présent si, pour la réalisation $x(t)$ observée sur l'intervalle $(-T, T)$, on a

$$\left(\int_{-T}^{T} x(t)\cos\omega_1 t\,dt\right)^2+\left(\int_{-T}^{T} x(t)\sin\omega_1 t\,dt\right)^2 \geqslant$$

$$\geqslant\left(\int_{-T}^{T} x(t)\cos\omega_0 t\,dt\right)^2+\left(\int_{-T}^{T} x(t)\sin\omega_0 t\,dt\right)^2. \qquad (5.199)$$

Le récepteur parfait de discrimination de deux signaux harmoniques correspondant au test (5.199) se compose de filtres adaptés à ces signaux (de durée finie, de pulsations ω_1 et ω_0; cf. problème 4.6), suivis de dispositifs d'élévation au carré, de sommation et de comparaison.

Notons que dans (5.199), on compare les carrés des valeurs des enveloppes du processus observé à la sortie de deux filtres adaptés aux signaux. C'est pourquoi un récepteur optimal de discrimination de deux signaux peut être interprété à l'aide de deux filtres adaptés aux signaux, de détecteurs linéaires séparant l'enveloppe des processus et d'un dispositif de comparaison.

Cherchons les probabilités d'erreurs. Pour l'exemple envisagé on trouve à partir de (5.195) à (5.196″)

$$m_{10}=m_{01}=d_T^2\,\frac{\sin 2\Delta T}{2\Delta T}, \qquad (5.200)$$

$$m_{11}=m_{00}=M_{21}=M_{20}=d_T^2, \qquad (5.201)$$

où

$$d_T^2=\frac{2A_0^2T}{N_0}, \qquad (5.202)$$

c'est-à-dire que d_T^2 est égal au rapport de l'énergie du signal à la densité spectrale du bruit.

Si $2T\Delta=(\omega_1-\omega_0)\,T=k\pi$, $k=1, 2, \ldots$, les signaux sont orthogonaux sur l'intervalle $(-T, T)$. En vertu de (5.200) les valeurs moyennes des composantes de vecteurs aléatoires normaux s'annulent. Ceci signifie que les variables aléatoires r_{T1} et r_{T0} sont alors réparties suivant une *loi de Rayleigh*. On peut alors obtenir l'expression des probabilités des erreurs sous forme fermée. En effet, la probabilité conditionnelle de prendre la décision que c'est le signal $s_1(t)$ qui a été transmis, alors qu'en réalité a été transmis le signal $s_0(t)$, est égale à

$$\alpha=P\{\gamma_1\mid H_0\}=P\{r_{T1}\geqslant r_{T0}\mid H_0\}=$$

$$=\frac{1}{d_T^4}\int_0^{\infty} x\exp\left(-\frac{x^2+d_T^4}{2d_T^2}\right)I_0(x)\int_x^{\infty} y e^{-\frac{y^2}{2d_T^2}}\,dy\,dx=$$

$$=\frac{1}{d_T^2}\int_0^\infty x\exp\left(-\frac{2x^2+d_T^4}{2d_T^2}\right)I_0(x)\,dx=$$

$$=\frac{1}{2}e^{-\frac{d_T^2}{4}}\int_0^\infty z\sqrt{2}\exp\left[-\frac{(z\sqrt{2})^2+\left(\frac{d_T}{\sqrt{2}}\right)^2}{2}\right]\times$$

$$\times I_0\left(z\sqrt{2}\frac{d_T}{\sqrt{2}}\right)d\left(z\sqrt{2}\right),$$

et comme la dernière intégrale est égale à l'unité on a

$$\alpha=\frac{1}{2}e^{-\frac{d_T^2}{4}}. \qquad (5.203)$$

Il est facile de voir que les probabilités conditionnelles d'erreurs de première et de seconde espèce sont égales entre elles, en effet

$$\beta=P\{\gamma_0\mid H_1\}=P\{r_{T0}>r_{T1}\mid H_1\}=$$

$$=\frac{1}{d_T^4}\int_0^\infty ye^{-\frac{y^2+d_T^4}{2d_T^2}}I_0(y)\int_y^\infty xe^{-\frac{x^2}{2d_T^2}}\,dx\,dy=\alpha.$$

Notons que l'algorithme optimal de discrimination de deux signaux à bande étroite $A\cos(\omega_0 t+\varphi_0)$ et $A\cos(\omega_1 t+\varphi_1)$ d'*amplitude inconnue* A et de phases aléatoires uniformément réparties ne diffère pas de (5.197). La grandeur A n'aura d'influence que sur le paramètre d_T^2 [cf. (5.202)] donnant la probabilité d'erreur.

5.4.5. Discrimination de plusieurs signaux. En continuant à à généraliser *) nous allons passer à l'étude du problème de la discrimination de plusieurs signaux. Supposons que l'on sache que le processus observé est un mélange additif d'un bruit normal stationnaire et de l'un des $m+1$ signaux déterministes $s_0(t)$, $s_1(t)$, ..., $s_m(t)$. L'ensemble des décisions contient maintenant non pas deux mais $m+1$ décisions γ_0, γ_1, . . ., γ_m, où γ_k est la décision que le processus observé se compose du bruit et du signal $s_k(t)$. Nous allons adopter le critère de qualité de Bayes, dans l'hypothèse supplémentaire que les coûts des décisions correctes sont nuls, et ceux des décisions erronées sont égaux. Supposons de plus que la transmission de chacun des signaux soit équiprobable **). En vertu de

*) Les résultats des paragraphes précédents de ce chapitre peuvent être obtenus à partir de ceux qui vont être donnés ci-dessous en posant $m=1$.

**) Si la probabilité a priori du signal $s_k(t)$ est égale à p_k, les résultats qui suivent restent vrais après substitution de $l_k(x_1, \ldots, x_N)$ à $\frac{p_k}{p_0}l_k(x_1, \ldots, x_N)$, $k=1, \ldots, m$.

§ 1.4.8 on peut alors formuler comme suit le test optimal de discrimination des signaux : c'est le signal $s_k(t)$, $k = 1, \ldots, m$ qui est présent si pour tous les $j \neq k$ pour l'échantillon observé $x_1, \ldots, x_N$ on a

$$\ln l_k(x_1, \ldots, x_N) \geqslant \ln l_j(x_1, \ldots, x_N) \qquad (5.204)$$

et

$$\ln l_k(x_1, \ldots, x_N) \geqslant 0 \qquad (5.204')$$

pour k donné, et le signal $s_0(t)$ si pour tous les k ($k = 1, \ldots, m$) on a

$$\ln l_k(x_1, \ldots, x_N) < 0. \qquad (5.204'')$$

Dans les inégalités mentionnées le rapport de vraisemblance est

$$l_k(x_1, \ldots, x_N) = \frac{W_N(x_1, \ldots, x_N \mid s_k)}{W_N(x_1, \ldots, x_N \mid s_0)}, \quad k = 1, \ldots, m. \qquad (5.205)$$

En prenant pour coordonnées observées les variables x_i données par (5.149), et en se limitant aux N premières coordonnées, le logarithme du k-ième rapport de vraisemblance s'écrit [cf. (5.151)]

$$\ln l_k(x_1, \ldots, x_N) = \frac{1}{2} \sum_{i=1}^{N} (a_{i0}^2 - a_{ik}^2) + \sum_{i=1}^{N} (a_{ik} - a_{i0}) x_i, \quad k = 1, \ldots, m, \qquad (5.206)$$

où

$$a_{ik} = \sqrt{\lambda_i} \int_{-T}^{T} s_k(t) \varphi_i(t) dt, \quad k = 0, 1, \ldots, m, \qquad (5.207)$$

et λ_i, $\varphi_i(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale (5.150) dont le noyau coïncide avec la fonction de corrélation du bruit.

Les relations (5.204) à (5.204″) et (5.206) permettent de trouver le test optimal de discrimination : c'est le signal $s_k(t)$ qui est présent (décision γ_k adoptée) si pour tous les $j \neq k$ ($j = 0, 1, \ldots, m$) on a

$$\sum_{i=1}^{N} (a_{ik} - a_{ij}) x_i \geqslant \frac{1}{2} \sum_{i=1}^{N} (a_{ik}^2 - a_{ij}^2). \qquad (5.208)$$

Si les résultats des observations ne peuvent pas être échantillonnés et si toute la réalisation $x(t)$ est utilisée sur l'intervalle d'observation, on obtient à partir de (5.208) pour $N \to \infty$ le test suivant [cf. (5.158)] : c'est le signal $s_k(t)$ qui est présent si pour tous les

$j \neq k$ $(j = 0, 1, \ldots, m)$ on a

$$\int_{-T}^{T} V_{kj}(t)\, x(t)\, dt \geqslant \frac{1}{2} \int_{-T}^{T} V_{kj}(t)\, [s_k(t) + s_j(t)]\, dt, \qquad (5.209)$$

où $V_{kj}(t)$ est la solution de l'équation intégrale linéaire non homogène

$$\int_{-T}^{T} B(t-u)\, V_{kj}(u)\, du = s_k(t) - s_j(t), \quad |t| \leqslant T. \qquad (5.210)$$

Lorsque l'on discerne des signaux noyés dans un bruit blanc additif, la solution de l'équation intégrale (5.210) est

$$V_{kj}(t) = \frac{1}{N_0} [s_k(t) - s_j(t)],$$

où N_0 est la densité spectrale du bruit. Dans ce cas (5.209) s'écrit

$$\int_{-T}^{T} s_k(t)\, x(t)\, dt \geqslant \int_{-T}^{T} s_j(t)\, x(t)\, dt + \frac{E_k - E_j}{2}, \quad j \neq k, \qquad (5.211)$$

où E_k et E_j sont les énergies des signaux $s_k(t)$ et $s_j(t)$ sur l'intervalle d'observation. L'inégalité (5.211) équivaut à [cf. (5.162')]

$$\int_{-T}^{T} [s_j(t) - x(t)]^2\, dt \geqslant \int_{-T}^{T} [s_k(t) - x(t)]^2\, dt, \quad j \neq k, \qquad (5.212)$$

celle-ci pouvant être interprétée comme suit: on adopte la décision sur la présence de celui des signaux $s_k(t)$ pour lequel la *distance* jusqu'à la réalisation observée est la moindre.

L'inégalité (5.211) montre que le récepteur optimal permettant de discriminer $m + 1$ signaux déterministes noyés dans un bruit blanc additif normal se compose de $m + 1$ filtres adaptés de fonctions de transition impulsionnelles

$$h_j^*(\tau) = s_j(T - \tau), \quad |\tau| \leqslant T, \qquad (5.213)$$

$$j = 0, 1, \ldots, m,$$

suivis de dispositifs de sélection et de comparaison. Pour les signaux d'énergie égale on compare les valeurs à la sortie des filtres adaptés et l'on prend la décision sur la présence de celui des signaux pour lequel cette valeur est maximale (fig. 5.4).

Nous allons également étudier le problème de la discrimination de $m + 1$ signaux quasi déterministes à bande étroite noyés dans un bruit normal additif

$$s_k(t) = a_k(t) \cos[\omega_k t - \psi_{sk}(t) + \varphi_k] \qquad (5.214)$$

si l'on connaît exactement $a_k(t)$ et $\psi_{sk}(t)$, les phases initiales φ_k étant aléatoires, indépendantes et uniformément réparties sur l'intervalle $(-\pi, \pi)$.

Soit ω_0 la plus petite des pulsations de signaux et soit

$$\Delta_k = \frac{\omega_k - \omega_0}{2} \ll \omega_0.$$

Choisissons pour pulsations centrales les grandeurs

$$\omega_k^* = \frac{\omega_k + \omega_0}{2}$$

en représentant en fonction de l'enveloppe et de la phase les processus à bande étroite recelant les signaux $s_k(t)$ et $s_0(t)$.

Fig. 5.4. Schéma d'un dispositif optimal de discrimination de $m+1$ signaux déterministes

D'une manière analogue à (5.189), écrivons l'expression du logarithme du k-ième rapport des fonctions de vraisemblance centrées:

$$\ln \Lambda_k (z_1, \ldots, z_N) = \ln I_0 (r_{Nk}) - \ln I_0 (r_{N0}) - \\ - \frac{1}{2} \sum_{i=1}^{N} [\,|s_{ik}|^2 - |s_{i0}|^2], \quad k = 1, \ldots, m, \tag{5.215}$$

où pour tout k de 0 à m on a

$$r_{Nk} = \left| \sum_{i=1}^{N} z_i s_{ik} \right| ; \tag{5.216}$$

$$s_{ik} = \sqrt{\lambda_i} \int_{-T}^{T} z_{sk}(t)\, e^{it\Delta_k} \overline{\varphi_k(t)}\, dt ; \tag{5.217}$$

$$z_{sk}(t) = a_k(t)\, e^{-i\psi_{sk}(t)}, \tag{5.218}$$

λ_i, $\varphi_i(t)$ étant comme auparavant les nombres caractéristiques et les fonctions propres de l'équation intégrale linéaire homogène, dont le noyau coïncide avec la fonction de corrélation $B_z(\tau)$ de l'enveloppe complexe du bruit [cf. (5.35')].

A partir de (5.204) à (5.204") et (5.215) on obtient le test suivant: c'est le signal $s_k(t)$ qui est présent (décision γ_k) si pour l'échantillon observé $z_1, \ldots, z_N$ et pour tous les $j \neq k$ on a

$$\ln I_0 (r_{Nk}) - \ln I_0 (r_{Nj}) \geqslant \frac{1}{2} \sum_{i=1}^{N} [\,|s_{ik}|^2 - |s_{jk}|^2]. \tag{5.219}$$

Si l'on effectue la discrimination d'après la réalisation $z(t)$ de l'enveloppe complexe sur l'intervalle $(-T, T)$, pour $N \to \infty$ on a en

vertu de (5.219)

$$\ln I_0(r_{Tk}) - \ln I_0(r_{Tj}) \geqslant \frac{d_{Tk}^2 - d_{Tj}^2}{2}, \quad j \neq k, \tag{5.220}$$

où pour j quelconque de 0 à m on a

$$r_{Tj} = \left| \int_{-T}^{T} U_j(t)\, \overline{z(t)}\, dt \right|, \tag{5.221}$$

$$d_{Tj}^2 = \int_{-T}^{T} U_j(t)\, \overline{z_{sj}(t)}\, e^{-it\Delta_j}\, dt, \tag{5.222}$$

$U_j(t)$ étant la solution de l'équation intégrale linéaire non homogène

$$\int_{-T}^{T} B_z(t-y)\, U_j(y)\, dy = z_{sj}(t)\, e^{it\Delta_j}, \quad |t| \leqslant T. \tag{5.223}$$

Lorsqu'on discerne les signaux noyés dans un bruit blanc additif de densité spectrale N_0 la solution de cette équation est

$$U_j(t) = \frac{1}{N_0} z_{sj}(t)\, e^{it\Delta_j}, \tag{5.224}$$

les grandeurs r_{Tj} et d_{Tj}^2 étant égales à

$$r_{Tj} = \frac{1}{N_0} \left| \int_{-T}^{T} \overline{z(t)}\, z_{sj}(t)\, e^{it\Delta_j}\, dt \right|, \tag{5.225}$$

$$d_{Tj}^2 = \frac{1}{N_0} \int_{-T}^{T} |z_{sj}(t)|^2\, dt = \frac{E_j}{N_0}, \tag{5.226}$$

où E_j est l'énergie du j-ième signal.

Lorsqu'on discerne des signaux de même énergie $E_j = E$, le test tenant compte du fait que la fonction $\ln I_0(r)$ est monotone peut se formuler comme suit : c'est le signal $s_k(t)$ qui est présent si pour tous les $j \neq k$ on a

$$r_{Tk} \geqslant r_{Tj} \tag{5.227}$$

ou

$$\left| \int_{-T}^{T} \overline{z(t)}\, z_{sk}(t)\, e^{it\Delta_k}\, dt \right| \leqslant \left| \int_{-T}^{T} \overline{z(t)}\, z_{sj}(t)\, e^{it\Delta_j}\, dt \right|. \tag{5.227'}$$

Comme les composantes en quadrature $z_{sj}(t)$ et $z(t)$ sont des fonctions variant lentement (cf. § 6.2.2, tome I), négligeant dans (5.227') les intégrales des produits de ces fonctions par les porteuses haute fréquence du type $\cos \omega t$ pour $\omega \geqslant \omega_0$, on peut donner le test suivant de discrimination des signaux d'énergie égale noyés dans un bruit blanc additif : on adopte la décision que c'est le signal $s_k(t)$

qui est présent si pour la réalisation observée $x(t)$ pour tous les $j \neq k$ on a

$$\left(\int_{-T}^{T} x(t)\, a_k(t) \cos[\omega_k t - \psi_{ks}(t)]\, dt\right)^2 + \\ + \left(\int_{-T}^{T} x(t)\, a_k(t) \sin[\omega_k t - \psi_{ks}(t)]\, dt\right)^2 \geqslant \\ \geqslant \left(\int_{-T}^{T} x(t)\, a_j(t) \cos[\omega_j t - \psi_{js}(t)]\, dt\right)^2 + \\ + \left(\int_{-T}^{T} x(t)\, a_j(t) \sin[\omega_j t - \psi_{js}(t)]\, dt\right)^2. \quad (5.228)$$

Le récepteur optimal permettant la discrimination se compose de $2m + 2$ filtres adaptés aux signaux $s_k(t)$, $|t| \leqslant T$, $k = 0, 1, \ldots, m$ et aux signaux

$$s_k^*(t) = a_k(t) \sin[\omega_k t - \psi_k(t)], \quad (5.229)$$
$$|t| \leqslant T, \; k = 0, 1, \ldots, m$$

(qui sont les transformées de Hilbert de $s_k(t)$; cf. annexe VI), ainsi que de dispositifs d'élévation au carré, de sommation et de choix de la plus grande valeur (fig. 5.5).

Notons que dans (5.228) on compare les carrés des valeurs des enveloppes du processus observé à la sortie de deux filtres adaptés aux signaux $s_k(t)$ et $s_j(t)$. C'est pourquoi le récepteur optimal destiné à discerner $m + 1$ signaux peut se composer seulement de $m + 1$ filtres adaptés, dont les signaux de sortie sont démodulés et puis appliqués à un dispositif de comparaison, discriminant la plus grande des $m+1$ valeurs de l'enveloppe (fig. 5.6).

Il est facile ici de calculer la probabilité de confusion des signaux en imposant les conditions suivantes *d'orthogonalité*:
pour $j \neq k$

$$\int_{-T}^{T} s_k(t)\, s_j(t)\, dt = 0, \quad (5.230)$$

$$\int_{-T}^{T} s_k^*(t)\, s_j(t)\, dt = 0. \quad (5.230')$$

La variable r_{Tk} est alors répartie suivant une loi de Rayleigh de paramètre $d_T^2 = \frac{E}{N_0}$ si l'hypothèse H_k (le signal s_k est présent) n'est pas vérifiée et suivant une loi de Rayleigh généralisée de paramètres

(d_T^1, d_T^2) si l'hypothèse H_k est vérifiée. La probabilité p_{err} de confusion du k-ième signal avec le j-ième, c'est-à-dire d'adopter la déci-

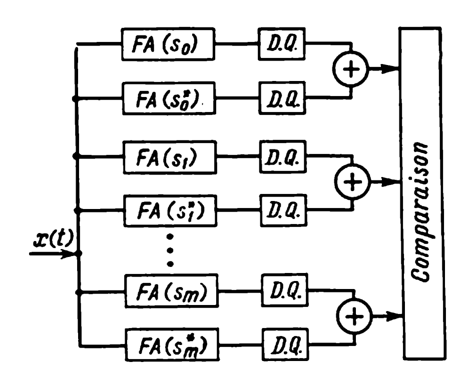


Fig. 5.5. Schéma d'un dispositif optimal de discrimination de $m + 1$ signaux quasi déterministes à bande étroite

Fig. 5.6. Schéma d'un dispositif optimal de discrimination équivalent à celui de la fig. 5.5

sion que c'est le signal $s_j(t)$ qui est présent alors qu'en réalité c'est le signal $s_k(t)$ qui a été envoyé est, pour tous k et j ($k \neq j$), égal à [cf. (5.203)]

$$p_{err} = \frac{m}{2} e^{-\frac{d_T^2}{4}}. \tag{5.231}$$

Problèmes

5.1. Généraliser les résultats du § 5.2.2 à la détection d'un signal quasi déterministe

$$s(t) = Aa(t) \cos[\omega_0 t - \psi_s(t) + \varphi_0)$$

d'amplitude A et de phase φ_0 aléatoires. Etudier le cas où la phase ne dépend pas de l'amplitude et est uniformément répartie. Montrer que pour une classe étendue de densités de probabilité $w_1(A)$ le test optimal de Bayes, permettant la détection du signal $s(t)$ noyé dans un bruit normal additif, ne dépend pas de $w_1(A)$ et est donné par une inégalité de la forme (5.44') ou (5.62'). Montrer, en particulier, que si l'amplitude suit une loi de Rayleigh, à condition que $M_2\{A\} = 2$, la probabilité de perte du signal est égale à

$$\beta(A) = e^{-\frac{A^2 d_T^2}{2}} \int_0^K x e^{-\frac{x^2}{2}} I_0(xA\,d_T)\,dx, \tag{1}$$

où le seuil K est déterminé par la probabilité donnée de fausse alarme

$$K = \sqrt{2 \ln \frac{1}{\alpha}}, \tag{1'}$$

d_T étant donné par (5.61). Pour un échantillon discret, d_T est remplacé par d_N [cf. (5.49)].

5.2. Pour les conditions du problème 5.1, dans le cas où l'amplitude A est répartie suivant une loi de Rayleigh de variance $M_2\{A\} = 2$, trouver le logarithme du rapport de vraisemblance centré

$$\ln \Lambda(z_1, \ldots, z_N) = -\ln(1+d_T^2) + \frac{r_T^2}{2(1+d_T^2)}. \tag{2}$$

Montrer que la probabilité de détection correcte et la probabilité de fausse alarme sont dans ce cas liées par la relation

$$1-\beta = \alpha^{\frac{1}{1+d_T^2}} \tag{3}$$

Dans les formules (2) et (3) les grandeurs r_T et d_T sont données par (5.60) et (5.61) respectivement. Pour des échantillons discrets, r_T et d_T sont remplacés par r_N et d_N [cf. (5.42) et (5.49)].

5.3. Montrer que la densité de probabilité de la somme des carrés de N variables aléatoires indépendantes, réparties chacune suivant une loi de Rayleigh généralisée

$$w_1(r) = r e^{-\frac{r^2+a^2}{2}} I_0(ar), \quad r > 0,$$

est de la forme

$$W_1(y) = \frac{1}{2}\left(\frac{y}{a^2}\right)^{\frac{N-1}{2}} e^{-\frac{y+a^2}{2}} I_{N-1}(a\sqrt{y}), \quad y > 0. \tag{4}$$

Obtenir à partir de (4) le cas particulier pour $a = 0$ de la densité de probabilité de la somme de N variables aléatoires indépendantes réparties chacune suivant une loi de Rayleigh (loi du χ^2 à $2N$ degrés de liberté), soit:

$$W_1(y) = \frac{1}{2^N (N-1)!} y^{N-1} e^{-y/2}, \quad y > 0. \tag{4'}$$

5.4. En utilisant les résultats du problème 5.3, trouver les probabilités de fausse alarme α et de perte d'un signal β lors de la détection d'un signal sinusoïdal à l'aide de l'algorithme (5.101) (le rapport de l'amplitude du signal à la valeur quadratique moyenne du bruit étant petit). Montrer que pour N quelconque on a

$$\alpha = 1 - \frac{\Gamma\left(N, \frac{2K_N}{a^2}\right)}{\Gamma(N)}, \tag{5}$$

$\Gamma(N, z)$ étant une fonction gamma incomplète [comparer avec (5.90)], soit:

$$\beta = \int_0^{\frac{2\sqrt{K_N}}{a}} x\left(\frac{x}{a}\right)^{N-1} e^{-\frac{x^2+a^2}{2}} I_{N-1}(ax)\,dx =$$

$$= \int_0^{\frac{2\sqrt{K_N}}{a}} x e^{-\frac{x^2+a^2}{2}} I_0(ax)\,dx + e^{-\left(\frac{a^2}{2}+\frac{2K_N}{a^2}\right)} \times$$

$$\times \sum_{k=1}^{N-1} \left(\frac{2\sqrt{K_N}}{a^2}\right)^k I_k(2\sqrt{K_N}), \tag{5'}$$

la dernière intégrale étant la fonction de répartition de la loi de Rayleigh généralisée [24].

Etudier le cas du critère du maximum de vraisemblance pour lequel on a

$$K_N = \frac{Na^2}{2}.$$

5.5. Montrer que la probabilité de détection d'un signal sinusoïdal d'amplitude constante A_0 pour $\frac{A_0}{\sigma} \gg 1$, lorsqu'on utilise l'algorithme (5.128') et le critère de Neumann-Pearson, peut être donnée par la formule

$$1-\beta = \frac{1}{\Gamma\left(\frac{N-1}{2}\right)} \Gamma\left[\frac{N-1}{2}, \sqrt{\frac{2}{N-1}} \left(\frac{A_0}{\sigma}\right)^2 (N-K)\right], \tag{6}$$

où $\Gamma(m, \lambda)$ est une fonction gamma incomplète et le seuil K est donné par la probabilité de fausse alarme.

5.6. Supposons que l'on vérifie l'hypothèse H_0 que la réalisation $x(t)$, $|t| \leqslant T$ appartient à un signal stochastique de fonction de corrélation $B_0(u, v)$ contre l'alternative H_1 que cette réalisation appartient à un signal stochastique de fonction de corrélation $B_1(u, v)$. Les signaux sont répartis normalement. Montrer que le dispositif optimal de discrimination des deux signaux stochastiques mentionnés est le même que celui qui est donné sur la figure 5.3, et que la fonction de transition impulsionnelle $h(u, v)$ d'un filtre linéaire de paramètres variables est donnée par l'équation intégrale suivante (cf. § 3.5.7 ainsi que [21]):

$$\int_{-T}^{T}\int_{-T}^{T} B_0(t, u)\, B_1(v, y)\, h(u, v)\, du\, dv = B_1(t, y) - B_0(t, y), \tag{7}$$

$$|t| \leqslant T, \quad |y| \leqslant T.$$

5.7. Montrer que la probabilité d'erreur lors de la discrimination de $m+1$ signaux orthogonaux déterministes $s_k(t)$, $k = 0, \ldots, m$ d'énergie égale E noyés dans un bruit blanc additif d'intensité N_0 est égale à

$$p_{\text{err}} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} F^m\left(x + \sqrt{\frac{2E}{N_0}}\right) dx. \tag{8}$$

5.8. Remplacer dans le problème 5.7. la condition d'orthogonalité par la condition

$$\int_{-T}^{T} s_k(t)\, s_j(t)\, dt = \lambda E, \quad \lambda \leqslant 1, \tag{9}$$

et montrer que dans ce cas on a

$$p_{\text{err}} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} F^m\left(x + \sqrt{\frac{2E(2-\lambda)}{N_0}}\right) dx. \tag{10}$$

Montrer que dans (9) le coefficient temporel λ de corrélation mutuelle des signaux ne peut être inférieur à $-\frac{1}{m}$ (et, par conséquent, la valeur minimale possible de λ, correspondant à la probabilité minimale d'erreur, est égale à $-\frac{1}{m}$).

Chapitre 6

SÉPARATION DES SIGNAUX NOYÉS DANS UN BRUIT

6.1. ESTIMATIONS DU MAXIMUM DE VRAISEMBLANCE DES PARAMÈTRES INCONNUS D'UN SIGNAL

6.1.1. Généralités. Nous allons maintenant passer à l'exposé de certaines questions relatives à la *séparation* des signaux noyés dans un bruit. Comme nous l'avons déjà noté au début du chapitre précédent, ce problème apparaît lorsqu'a priori *on sait* que le signal est présent dans le processus aléatoire observé qui représente une combinaison du signal et du bruit, mais que restent *inconnues* certaines caractéristiques du signal contenant sous la forme codée l'information utile. Ces caractéristiques peuvent être des nombres inconnus ou des fonctions inconnues du temps, des variables aléatoires ou des processus aléatoires. Le problème de la séparation consiste à trouver les procédures, optimales vis-à-vis des critères de qualité adoptés, de construction des *estimations* des caractéristiques mentionnées d'après les résultats des observations, qui peuvent se présenter sous la forme d'un échantillon de taille finie ou d'une réalisation continue à l'entrée du récepteur *).

Tout comme dans le chapitre précédent, notre exposé se limite essentiellement au cas où les bruits sont additifs, normaux, stationnaires, statistiquement indépendants du signal. Nous allons commencer par le processus somme d'un bruit $\xi(t)$ de moyenne nulle et d'un signal déterministe $s(t;\ A, \varphi, \vartheta_1, \ldots, \vartheta_M)$ dépendant des paramètres inconnus $A, \varphi, \vartheta_1, \ldots, \vartheta_M$. Puis nous étudierons les estimations de Bayes des paramètres aléatoires d'un signal quasi déterministe, connaissant leur densité de probabilité conjointe a priori. Dans les §§ 6.1 et 6.2 on trouvera une étude détaillée et l'interprétation radiotechnique des résultats mathématiques mentionnés au § 3.6.

6.1.2. Estimations simultanées de l'amplitude et de la phase d'un signal harmonique. Au § 3.6.2. nous avons étudié en détail l'estimation de l'amplitude inconnue a d'un signal déterministe $as(t)$ noyé dans un bruit normal additif.

Nous y avons montré que l'estimation efficace non biaisée de l'amplitude inconnue s'obtient par intégration pondérée de la réalisation $x(t)$ du processus, les poids étant déterminés par la fonction

*) Notons que la procédure de construction de l'estimation des caractéristiques du signal peut être réunie avec la détection du signal (pour plus de détail voir [11]).

de corrélation du bruit. Dans le cas d'un bruit blanc cette estimation s'obtient à la sortie d'un filtre adapté au signal $s(t)$, et le rapport de la variance de l'estimation (minimale possible) au carré de l'amplitude estimée est alors égal au rapport de la densité spectrale du bruit à l'énergie du signal [cf. (3.177)].

Nous allons maintenant étudier le cas simple de l'estimation simultanée de deux paramètres inconnus, de l'amplitude A et de la phase φ d'un signal harmonique déterministe $s(t) = A\cos(\omega_0 t - \varphi)$ de pulsation ω_0 connue, noyé dans un bruit normal à bande étroite de fonction de corrélation $B(\tau)$ donnée. En utilisant la représentation complexe de la réalisation observée et du signal

$$x(t) = \operatorname{Re} z(t)\, e^{i\omega_0 t}, \tag{6.1}$$

$$s(t;\, A,\, \varphi) = \operatorname{Re} A e^{-i\varphi} e^{i\omega_0 t}, \tag{6.1'}$$

on peut écrire le système d'équations *du maximum de vraisemblance* donnant les estimations correspondant à ce critère. En vertu de (3.161′) pour $V(t;\, A,\, \varphi) = Ae^{-i\varphi}\, U(t)$ on a

$$e^{-i\varphi} \int_{-T}^{T} U(t)\overline{z(t)}e^{-i\omega_0 t}\,dt - A\int_{-T}^{T} U(t)e^{-i\omega_0 t}\,dt +$$

$$+ e^{i\varphi} \int_{-T}^{T} \overline{U(t)}\, z(t)\, e^{i\omega_0 t}\,dt - A\int_{-T}^{T} \overline{U(t)}\, e^{i\omega_0 t}\,dt = 0, \tag{6.2}$$

$$e^{-i\varphi} \int_{-T}^{T} U(t)\overline{z(t)}e^{-i\omega_0 t}\,dt - e^{i\varphi}\int_{-T}^{T} \overline{U(t)}\, z(t)\, e^{i\omega_0 t}\,dt = 0, \tag{6.2'}$$

où $U(t)$ est la solution de l'équation intégrale

$$\int_{-T}^{T} B_z(t-y)\, U(y)\, dy = e^{i\omega_0 t}, \quad |t| \leqslant T. \tag{6.3}$$

En ajoutant (6.2) et (6.2′) on obtient l'équation suivante reliant $\hat{A}$ et $\hat{\varphi}$:

$$\hat{A}e^{i\hat{\varphi}} = \frac{\int_{-T}^{T} U(t)\overline{z(t)}\, e^{-i\omega_0 t}\,dt}{\operatorname{Re}\int_{-T}^{T} U(t)\, e^{-i\omega_0 t}\,dt}. \tag{6.4}$$

A partir de (6.4) on peut obtenir directement les formules donnant l'estimation de l'amplitude et de la phase inconnues, on a

$$\hat{A} = \frac{\left| \int_{-T}^{T} U(t)\, \overline{z(t)}\, e^{-i\omega_0 t}\, dt \right|}{\operatorname{Re} \int_{-T}^{T} U(t)\, e^{-i\omega_0 t}\, dt}, \tag{6.5}$$

$$\hat{\varphi} = \operatorname{arc\,tg} \frac{\operatorname{Im} Y}{\operatorname{Re} Y}, \tag{6.5'}$$

où

$$Y = \int_{-T}^{T} \overline{U(t)}\, z(t)\, e^{-i\omega_0 t}\, dt. \tag{6.5''}$$

Si le bruit additif est un bruit *blanc* de densité spectrale N_0, on a $U(t) = \frac{1}{N_0} e^{i\omega_0 t}$ et on obtient à partir de (6.4)

$$\hat{A} e^{i\hat{\varphi}} = \frac{1}{2T} \int_{-T}^{T} z(t)\, dt. \tag{6.6}$$

A partir de (6.6) on obtient l'estimation suivante pour la phase:

$$\hat{\varphi} = \operatorname{arc\,tg} \frac{\operatorname{Im} \int_{-T}^{T} z(t)\, dt}{\operatorname{Re} \int_{-T}^{T} z(t)\, dt} = \operatorname{arc\,tg} \frac{\int_{-T}^{T} C(t)\, dt}{\int_{-T}^{T} A(t)\, dt}, \tag{6.7}$$

où $A(t)$ et $C(t)$ sont les composantes en quadrature de la réalisation observée (cf. § 5.2.2). Comme ces composantes sont des processus variant lentement par rapport à $\cos \omega_0 t$, on peut négliger les intégrales des produits $A(t)$ et $C(t)$ par les fonctions haute fréquence $\cos \omega t$ pour $\omega \geqslant \omega_0$. L'estimation (6.7) peut, suivant la réalisation observée $x(t)$, s'écrire comme suit:

$$\hat{\varphi} = \operatorname{arc\,tg} \frac{\int_{-T}^{T} x(t) \sin \omega_0 t\, dt}{\int_{-T}^{T} x(t) \cos \omega_0 t\, dt}. \tag{6.8}$$

Dans ce cas l'estimation de l'amplitude est égale à

$$\hat{A} = \left| \frac{1}{2T} \int_{-T}^{T} z(t)\, dt \right| = \frac{1}{2T} \sqrt{\left(\int_{-T}^{T} A(t)\, dt \right)^2 + \left(\int_{-T}^{T} C(t)\, dt \right)^2} \tag{6.9}$$

ou, compte tenu de la remarque concernant (6.8), on trouve définitivement

$$\hat{A}=\left[\left(\frac{1}{2T}\int_{-T}^{T} x(t)\cos\omega_0 t\,dt\right)^2+\left(\frac{1}{2T}\int_{-T}^{T} x(t)\sin\omega_0 t\,dt\right)^2\right]^{\frac{1}{2}}. \quad (6.10)$$

En comparant (6.9), (6.10) et (5.67), (5.67′) on voit que le dispositif de détection d'un signal harmonique quasi déterministe noyé dans un bruit blanc peut être utilisé directement pour l'estimation de l'amplitude du signal, et ses différents éléments (intégrateur des composantes en quadratures) pour l'estimation de la phase [cf. (6.7) et (6.8)].

Nous allons maintenant étudier les propriétés essentielles des estimations de l'amplitude et de la phase d'un signal harmonique noyé dans un bruit blanc. Notons avant tout que les intégrales $\frac{1}{2T}\int_{-T}^{T} A(t)\,dt$ et $\frac{1}{2T}\int_{-T}^{T} C(t)\,dt$ sont des variables aléatoires indépendantes réparties suivant une loi normale de paramètres $\left(A\cos\varphi, \frac{N_0}{2T}\right)$ et $\left(A\sin\varphi, \frac{N_0}{2T}\right)$ respectivement. Ainsi, en vertu de (6.7) et (6.9) les estimations $\hat{A}$ et $\hat{\varphi}$ sont le *module* et la *phase* d'un vecteur plan à composantes indépendantes normalement réparties et ayant les mêmes variances. Par conséquent, l'estimation $\hat{A}$ de l'amplitude A [cf. (6.9)] est répartie suivant une loi de Rayleigh généralisée de paramètres $\left(A, \frac{N_0}{2T}\right)$, qui pour $d_T \to \infty$ *tend asymptotiquement vers la loi normale.* La moyenne de l'estimation $\hat{A}$ est égale à [cf. (3.74), tome I]

$$m_1\{\hat{A}\}=\frac{A}{d_T}\sqrt{\frac{\pi}{2}}\left[\left(1+\frac{d_T^2}{2}\right)I_0\left(\frac{d_T^2}{4}\right)+\frac{d_T^2}{2}I_1\left(\frac{d_T^2}{4}\right)\right]e^{-\frac{d_T^2}{4}}, \quad (6.11)$$

où d_T^2 est le rapport de l'énergie du signal à la densité spectrale du bruit, c'est-à-dire

$$d_T^2=\frac{2TA^2}{N_0}. \quad (6.11')$$

Ainsi, l'estimation du maximum de vraisemblance de l'amplitude se trouve *biaisée.*

Quand $d_T^2 \gg 1$ on a [cf. (3.77), tome I]

$$m_1\{\hat{A}\}\sim\left(1+\frac{1}{2d_T^2}\right), \quad (6.12)$$

et, par conséquent, pour $d_T \to \infty$, $\hat{A}$ est une estimation *asymptotiquement correcte* (non biaisée) de l'amplitude du signal.

La variance de l'estimation $\hat{A}$ est égale à [cf. (3.75), tome I]

$$M_2\{\hat{A}\} = A^2\left(1+\frac{2}{d_T^2}\right) - m_1^2\{\hat{A}\}. \tag{6.13}$$

Quand $d_T^2 \gg 1$ on a [cf. (3.78), tome I]

$$M_2\{\hat{A}\} \sim \frac{A^2}{d_T^2}\left(1-\frac{1}{4d_T^2}\right). \tag{6.14}$$

Il est facile de calculer la densité de probabilité de l'estimation du maximum de vraisemblance de la phase $\hat{\varphi}$ (car cette estimation, comme nous l'avons noté ci-dessus, est la phase d'un vecteur de composantes normales et indépendantes), on a

$$\hat{\varphi} = \operatorname{arc\,tg} \frac{\int\limits_{-T}^{T} C(t)\,dt}{\int\limits_{-T}^{T} A(t)\,dt}. \tag{6.15}$$

En vertu de (3.81) du premier tome, la densité de probabilité de $\hat{\varphi}$ est égale à

$$w_1(\vartheta) = \frac{1}{2\pi} e^{-\frac{d_T^2}{2}} \left\{1 + d_T\sqrt{2\pi}\cos(\vartheta-\varphi)\, e^{\frac{d_T^2\cos^2(\vartheta-\varphi)}{2}} \times \right.$$
$$\left. \times F[d_T\cos(\vartheta-\varphi)]\right\}, \quad |\vartheta-\varphi| \leqslant \pi. \tag{6.16}$$

Comme $w_1(\vartheta)$ est une fonction paire par rapport à $\vartheta - \varphi$, on en conclut que l'estimation du maximum de vraisemblance est une estimation non biaisée de la phase inconnue d'un signal harmonique noyé dans un bruit normal additif.

La variance de l'estimation $\hat{\varphi}$ est égale à [cf. (3.93), tome I]

$$M_2\{\hat{\varphi}\} = \frac{\pi^2}{3} + 4\pi \sum_{n=1}^{\infty} (-1)^n \frac{a_n}{n^2}, \tag{6.17}$$

où

$$a_n = \frac{\Gamma\left(1+\frac{n}{2}\right) d_T^n}{\pi n!\, 2^{n/2}}\, {}_1F_1\left(\frac{n}{2},\ n+1,\ -\frac{d_T^2}{2}\right). \tag{6.17'}$$

Quand $d_T^2 \gg 1$, on voit de (6.16) que la répartition de $\hat{\varphi}$ tend asymptotiquement vers une loi normale de moyenne nulle et de variance égale à [cf. (3.88), tome I]

$$M_2\{\hat{\varphi}\} \sim \frac{1}{d_T^2}. \tag{6.18}$$

Ayant trouvé la matrice d'information relative aux paramètres estimés A et φ, on peut montrer que les estimations $\hat{A}$ et $\hat{\varphi}$ du maximum de vraisemblance sont asymptotiquement simultanément efficaces (cf. problème 6.3).

Ainsi, conformément à la théorie générale (cf. § 2.3.1) les estimations étudiées du maximum de vraisemblance de l'amplitude et de la phase d'un signal noyé dans un bruit blanc normal additif sont *asymptotiquement non biaisées, simultanément efficaces et normales*, l'estimation de la phase étant non biaisée (pour tout d_T fini) et celle de l'amplitude biaisée.

6.1.3. Estimations des paramètres d'un signal à bande étroite noyé dans un bruit blanc additif. Soit un signal à bande étroite

$$s(t;\ A,\ \varphi,\ \vartheta_1,\ \ldots,\ \vartheta_M) = A \operatorname{Re} a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)\, e^{-i\varphi} e^{i\omega_0 t}, \tag{6.19}$$

caractérisé, en plus de l'amplitude $A > 0$ et de la phase φ, par M paramètres inconnus de l'enveloppe complexe. Il y a lieu de trouver les estimations du maximum de vraisemblance de ces paramètres d'après la réalisation du processus $x(t) = \operatorname{Re} z(t)\, e^{i\omega_0 t}$ qui est un mélange additif du signal mentionné et d'un *bruit blanc normal* de densité spectrale N_0.

En utilisant (5.54) (avec des changements de notations évidents) et compte tenu de

$$V(t) = \frac{A}{N_0} a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)\, e^{-i\varphi},$$

on peut écrire comme suit le logarithme de la fonctionnelle du rapport de vraisemblance:

$$\ln l\,[z(t)\,|\,A,\ \varphi,\ \vartheta_1,\ \ldots,\ \vartheta_M] = \frac{2A}{N_0} \operatorname{Re} e^{-i\varphi} \int_{-T}^{T} a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)\, \overline{z(t)}\, dt - \\ - \frac{A^2}{N_0} \int_{-T}^{T} |a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)|^2\, dt. \tag{6.20}$$

Supposons que les signaux soient normés de telle sorte que pour $\vartheta_1,\ \ldots,\ \vartheta_M$ quelconques on a

$$\frac{1}{2T} \int_{-T}^{T} |a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)|^2\, dt = 1. \tag{6.21}$$

La condition (6.21) signifie que l'énergie du signal est constante sur l'intervalle d'observation, en effet

$$E = \int_{-T}^{T} s^2(t;\ A,\ \varphi,\ \vartheta_1,\ \ldots,\ \vartheta_M)\, dt = \\ = A^2 \int_{-T}^{T} |a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)|^2 \cos^2 [\omega_0 t + \\ + \arg a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M) - \varphi]\, dt \approx A^2 T,$$

où l'on a négligé l'intégrale des produits des fonctions variant lentement par la porteuse haute fréquence.

Compte tenu de (6.21) on peut écrire (6.20) comme suit

$$\ln l\,]z(t)\,|\,A,\ \varphi,\ \vartheta_1,\ \ldots,\ \vartheta_M[\ =$$

$$= \frac{1}{2TN_0}\left|\int_{-T}^{T}\overline{a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)}\,z(t)\,dt\right|^2 -$$

$$-\left|A\sqrt{\frac{2T}{N_0}}\,e^{i\varphi} - \frac{1}{\sqrt{2TN_0}}\int_{-T}^{T}\overline{a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)}\,z(t)\,dt\right|^2. \quad (6.22)$$

Soient $\hat{A}$, $\hat{\varphi}$, $\hat{\vartheta}_1, \ldots, \hat{\vartheta}_M$ les estimations du maximum de vraisemblance. Comme le premier terme de (6.22) est positif et ne dépend ni de A ni de φ, le maximum du logarithme de la fonctionnelle du rapport de vraisemblance par rapport aux paramètres A et φ a lieu lorsque le second terme (négatif) est nul. D'où la relation entre les estimations

$$\hat{A}e^{i\hat{\varphi}} = \frac{1}{2T}\int_{-T}^{T}\overline{\hat{a}(t;\ \vartheta_1,\ \ldots,\ \hat{\vartheta}_M)}\,z(t)\,dt, \quad (6.23)$$

et

$$\hat{A} = \left|\frac{1}{2T}\int_{-T}^{T}\overline{a(t;\ \hat{\vartheta}_1,\ \ldots,\ \hat{\vartheta}_M)}\,z(t)\,dt\right|, \quad (6.23')$$

$$\hat{\varphi} = \operatorname{arc\,tg}\frac{\operatorname{Im}\int_{-T}^{T}\overline{a(t;\ \hat{\vartheta}_1,\ \ldots,\ \hat{\vartheta}_M)\,z(t)}\,dt}{\operatorname{Re}\int_{-T}^{T}\overline{a(t;\ \hat{\vartheta}_1,\ \ldots,\ \hat{\vartheta}_M)\,z(t)}\,dt}. \quad (6.23'')$$

Notons que, comme on pouvait s'y attendre, pour $a(t) \equiv 1$ la formule (6.23) devient (6.6).

Il s'agit maintenant de trouver le maximum absolu du premier terme de (6.22), c'est-à-dire de la fonction suivante des paramètres $\vartheta_1, \ldots, \vartheta_M$:

$$Q(\vartheta_1,\ \ldots,\ \vartheta_M) = \frac{1}{2TN_0}\left|\int_{-T}^{T}\overline{a(t;\ \vartheta_1,\ \ldots,\ \vartheta_M)}\,z(t)\,dt\right|^2. \quad (6.24)$$

On sait que dans le problème envisagé de la *séparation du signal*, la réalisation observée est la somme d'un signal et d'un bruit. L'enveloppe complexe de cette réalisation est

$$z(t) = A_0\,a(t;\ \vartheta_1^0,\ \ldots,\ \vartheta_M^0)\,e^{-i\varphi_0} + z_1(t), \quad (6.25)$$

où $z_1(t)$ est un processus aléatoire normal stationnaire complexe à spectre uniforme, et les grandeurs A_0, φ_0, $\vartheta_1^0, \ldots, \vartheta_M^0$ sont les para-

mètres du signal qu'il y a lieu d'estimer d'après la réalisation $x(t) = \operatorname{Re} z(t)\, e^{i\omega_0 t}$.

En portant (6.25) dans (6.24) écrivons comme suit la fonction $Q(\vartheta_1, \ldots, \vartheta_M)$:

$$Q(\vartheta_1, \ldots, \vartheta_M) = \frac{2TA_0^2}{N_0}\,[h(\vartheta_1, \ldots, \vartheta_M) + 2\operatorname{Re}\overline{\Psi(\vartheta_1, \ldots, \vartheta_M)}\,\chi(\vartheta_1, \ldots, \vartheta_M)\, e^{i\varphi_0} + |\chi(\vartheta_1, \ldots, \vartheta_M)|^2], \tag{6.26}$$

avec les notations

$$h(\vartheta_1, \ldots, \vartheta_M) = |\Psi(\vartheta_1, \ldots, \vartheta_M)|^2 = \left|\frac{1}{2T}\int_{-T}^{T} a(t;\, \vartheta_1^0, \ldots, \vartheta_M^0)\,\overline{a(t;\, \vartheta_1, \ldots, \vartheta_M)}\, dt\right|^2; \tag{6.26'}$$

$$\chi(\vartheta_1, \ldots, \vartheta_M) = \frac{1}{2TA_0}\int_{-T}^{T} \overline{a(t;\, \vartheta_1, \ldots, \vartheta_M)}\, z_1(t)\, dt, \tag{6.26''}$$

h et Ψ étant des fonctions déterministes, et $\chi(\vartheta_1, \ldots, \vartheta_M)$ une variable aléatoire normale de moyenne nulle et de variance égale à l'inverse du rapport $d_T^2 = \frac{2TA_0^2}{N_0}$ de l'énergie du signal à la densité spectrale du bruit.

On appelle *fonction d'incertitude* pour des valeurs $\vartheta_1^0, \ldots, \vartheta_M^0$ données la fonction

$$\Psi(\vartheta_1, \ldots, \vartheta_M) = \frac{1}{2T}\int_{-T}^{T} a(t;\, \vartheta_1^0, \ldots, \vartheta_M^0)\,\overline{a(t;\, \vartheta_1, \ldots, \vartheta_M)}\, dt, \tag{6.27}$$

dépendant seulement de la forme de l'enveloppe complexe et de ses paramètres. En vertu de l'inégalité de Bouniakovski-Schwartz on a (cf. page 76, tome I)

$$|\Psi(\vartheta_1, \ldots, \vartheta_M)| \leqslant \left[\frac{1}{2T}\int_{-T}^{T} |a(t;\, \vartheta_1^0, \ldots, \vartheta_M^0)|^2\, dt \times \frac{1}{2T}\int_{-T}^{T} |a(t;\, \vartheta_1, \ldots, \vartheta_M)|^2\, dt\right]^{1/2}$$

ou, compte tenu de (6.21),

$$|\Psi(\vartheta_1, \ldots, \vartheta_M)| \leqslant 1. \tag{6.27'}$$

Pour $\vartheta_i = \vartheta_i^0$, $i = 1, \ldots, M$ la fonction $|\Psi|$ atteint son maximum égal à l'unité [cf. (6.27)]. Ceci correspond au cas où la composante de bruit dans (6.25) disparaît, c'est-à-dire quand $d_T \to \infty$.

On a ainsi

$$\Psi(\vartheta_1^0, \ldots, \vartheta_M^0) = |\Psi(\vartheta_1^0, \ldots, \vartheta_M^0)| = 1. \quad (6.27'')$$

La valeur moyenne prise sur l'ensemble des réalisations de la fonction Q étant

$$m_1\{Q(\vartheta_1, \ldots, \vartheta_M)\} = d_T^2\left[h(\vartheta_1, \ldots, \vartheta_M) + \frac{1}{d_T^2}\right]$$

et le premier terme entre crochets étant, en vertu de (6.27″), voisin de l'unité si

$$d_T^2 = \frac{2TA_0^2}{N_0} \gg 1, \quad (6.28)$$

en nous limitant dans ce qui suit à la condition (6.28), cherchons le maximum absolu de la fonction Q en négligeant le dernier terme dans (6.26). Cette fonction peut en général avoir de nombreux maxima dans l'espace à M dimensions des paramètres inconnus ϑ_i. Lorsque la réalisation ne contient pas de composante de bruit, comme nous l'avons déjà noté ci-dessus, cette fonction admet un maximum pour $\vartheta_i = \vartheta_i^0$, $i = 1, \ldots, M$. La présence du bruit donne un déplacement de ce maximum, mais si la condition (6.28) se trouve réalisée, on peut supposer ce déplacement *petit*, de sorte que les estimations $\hat{\vartheta}_i$ correspondant au maximum absolu de la fonction de vraisemblance sont *voisines* des valeurs réelles des paramètres estimés ϑ_i. Ceci dit, développons la fonction $Q(\vartheta_1, \ldots, \vartheta_M)$ de M variables en série de Taylor au voisinage du point $\vartheta_1^0, \ldots, \vartheta_M^0$. Nous ne conserverons que les termes du second ordre de petitesse et nous négligerons le dernier terme dans (6.26). Au point $(\vartheta_1^0, \ldots, \vartheta_M^0)$ on a

$$\frac{\partial h}{\partial \vartheta_k} = 0,$$

car en ce point la fonction $h(\vartheta_1, \ldots, \vartheta_M)$ est maximale. Compte tenu des remarques faites ci-dessus, on obtient à partir de (6.26)

$$Q(\vartheta_1, \ldots, \vartheta_M) = Q(\vartheta_1^0, \ldots, \vartheta_M^0) + \\ + \sum_{k=1}^{M} (\vartheta_k - \vartheta_k^0) \left.\frac{\partial Q(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k}\right|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0} + \\ + \frac{1}{2}\sum_{k=1}^{M}\sum_{j=1}^{M} (\vartheta_k - \vartheta_k^0)(\vartheta_j - \vartheta_j^0) \left.\frac{\partial^2 Q(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k \, \partial \vartheta_j}\right|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0},$$

où

$$Q(\vartheta_1^0, \ldots, \vartheta_M^0) = d_T^2\,[1 + 2\,\mathrm{Re}\,\chi(\vartheta_1^0, \ldots, \vartheta_M^0)\,e^{i\varphi_0}];$$

$$\left.\frac{\partial Q(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k}\right|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0} = 2d_T^2\,\mathrm{Re}\,e^{i\varphi_0} \times \\ \times \left[\frac{\partial \chi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k} + \chi(\vartheta_1, \ldots, \vartheta_M)\,\overline{\frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k}}\right]_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}.$$

En dérivant par rapport à ϑ_k le développement obtenu de la fonction Q et égalant à zéro le résultat obtenu, on trouve le système suivant d'équations de vraisemblance:

$$\sum_{j=1}^{M} Q_{kj}(\vartheta_j - \vartheta_j^0) = \zeta_k, \quad k = 1, \ldots, M, \tag{6.29}$$

où l'on a employé les notations

$$\zeta_k = -\frac{\partial Q(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k}\bigg|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}, \tag{6.29'}$$

$$Q_{kj} = \frac{\partial^2 Q(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k \, \partial \vartheta_j}\bigg|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}. \tag{6.29''}$$

Si le déterminant de la matrice Q composée d'éléments Q_{kj} est différent de zéro, la matrice inverse $\mathbf{Q}^{-1}$ existe, ses éléments étant égaux au rapport des cofacteurs au déterminant de la matrice Q. Désignons par q_{kj} les éléments de la matrice inverse, nous pouvons alors écrire comme suit la solution du système d'équations linéaires (6.29) par rapport à l'erreur:

$$\hat{\vartheta}_j - \vartheta_j^0 = \sum_{i=1}^{M} \zeta_k q_{kj}, \quad j = 1, \ldots, M, \tag{6.30}$$

où pour l'approximation adoptée [cf. (6.26) et (6.28)]

$$\zeta_k = 2d_T^2 \operatorname{Re} \left\{ e^{i\varphi_0} \frac{1}{2TA_0} \int_{-T}^{T} \left[\overline{a(t; \vartheta_1^0, \ldots, \vartheta_M^0)} \times \right.\right.$$

$$\times \left(\frac{1}{2T} \int_{-T}^{T} a(t; \vartheta_1^0, \ldots, \vartheta_M^0) \frac{\overline{\partial a(t; \vartheta_1^0, \ldots, \vartheta_M^0)}}{\partial \vartheta_k} dt \right) +$$

$$\left.\left. + \frac{\overline{\partial a(t; \vartheta_1^0, \ldots, \vartheta_M^0)}}{\partial \vartheta_k} \right] z_1(t)\, dt. \tag{6.30'}$$

Les grandeurs ζ_k et q_{kj} sont aléatoires et dépendent de la composante de bruit de la réalisation. Cependant, en négligeant toujours le dernier terme (quadratique par rapport à χ) dans (6.26), il ne reste qu'à conserver dans (6.29″) pour Q_{kj} l'apport de la composante déterministe de (6.26). Pour $d_T^2 \gg 1$ il faut poser

$$Q_{kj} = d_T^2 \frac{\partial^2 h(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_k \, \partial \vartheta_j}\bigg|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}, \tag{6.31}$$

c'est-à-dire que les éléments de la matrice Q sont égaux au produit de la grandeur d_T^2 et de la dérivée seconde mixte du carré du module de la fonction d'incertitude au point correspondant aux valeurs réelles des paramètres estimés.

Compte tenu de (6.31), on arrive à la conclusion que les erreurs $\hat{\vartheta}_j - \vartheta_j^0$ [cf. (6.30)] sont les sommes pondérées (avec des poids égaux aux éléments de la matrice inverse de Q) des variables aléatoires *normales* ζ_k de moyennes nulles et de covariances égales à

$$m_1\{\zeta_k\zeta_j\} = 2d_T^4 \operatorname{Re} m_1\Big\{\Big[\frac{\partial\chi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_k} + \chi(\vartheta_1, \ldots, \vartheta_M)\overline{\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_k}}\Big]\Big[\overline{\frac{\partial\chi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_j}} + \overline{\chi(\vartheta_1, \ldots, \vartheta_M)}\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_j}\Big]\Big\}_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}. \tag{6.32}$$

Dans la formule (6.32) on a tenu compte du fait que pour l'enveloppe complexe on a $m_1\ \{z_1(t)\, z_1(t')\} = 0$, quels que soient t et t' [cf. (5.37')].

Il s'ensuit tout d'abord qu'en vertu de la théorie générale, les estimations $\hat{\vartheta}_j$ sont, pour $d_T^2 \gg 1$, asymptotiquement *normales*. Les valeurs moyennes de ces estimations sont [cf. (6.30)]

$$m_1\ \{\hat{\vartheta}_j\} = \vartheta_j^0, \tag{6.33}$$

c'est-à-dire que les estimations du maximum de vraisemblance sont, pour $d_T^2 \gg 1$, asymptotiquement *non biaisées*. Les éléments de la matrice de corrélation des erreurs sont

$$m_1\{(\hat{\vartheta}_k - \vartheta_k^0)(\hat{\vartheta}_n - \vartheta_n^0)\} = \sum_{i=1}^{M}\sum_{j=1}^{M} q_{ki}q_{jn}m_1\{\zeta_i\zeta_j\}. \tag{6.34}$$

On peut trouver la covariance $m_1\ \{\zeta_i\zeta_j\}$ à partir de (6.32). En effet,

$$m_1\Big\{\overline{\chi(\vartheta_1^0, \ldots, \vartheta_M^0)}\frac{\partial\chi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_i}\Big\} = \frac{1}{d_T^2}\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_i}$$

et $m_1\ \{|\ \chi\ |^2\} = 1/d_T^2$, donc

$$m_1\{\zeta_i\zeta_j\} = 2d_T^2 \operatorname{Re}\Big\{\Big[\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_j}\overline{\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_i}} + \frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_j}\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_i} + \overline{\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_j}}\,\overline{\frac{\partial\Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_i}} + \frac{1}{2T}\int_{-T}^{T}\frac{\partial a(t;\ \vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_i}\overline{\frac{\partial a(t;\ \vartheta_1, \ldots, \vartheta_M)}{\partial\vartheta_j}}\,dt\Big]_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}\Big\} \tag{6.35}$$

De plus, en dérivant (6.21) et (6.27) par rapport à ϑ_i et ϑ_j on trouve (compte tenu de ce que $\left.\frac{\partial h}{\partial \vartheta_j}\right|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0} = 0$)

$$\operatorname{Re}\left\{\left.\frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_j}\right|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}\right\} = 0, \tag{6.36}$$

$$\operatorname{Re}\left\{\left[\frac{\partial^2 \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i \partial \vartheta_j} + \frac{1}{2T}\int_{-T}^{T} \frac{\partial a(t; \vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i} \times \right.\right.$$
$$\left.\left. \times \overline{\frac{\partial a(t; \vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_j}}\, dt\right]_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}\right\} = 0,$$
$$i, j = 1, \ldots, M. \tag{6.36'}$$

En vertu de (6.26′) on a

$$\frac{\partial^2 h(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i \partial \vartheta_j} = 2 \operatorname{Re}\left[\frac{\partial^2 \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i \partial \vartheta_j} + \right.$$
$$\left. + \frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i} \overline{\frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_j}}\right], \tag{6.37}$$

ce qui, compte tenu de (6.36′), donne

$$\left.\frac{\partial^2 h(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i \partial \vartheta_j}\right|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0} =$$
$$= 2 \operatorname{Re}\left\{\left[\frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i} \overline{\frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_j}} - \frac{1}{2T}\int_{-T}^{T} \frac{\partial a(t; \vartheta_1, \ldots \vartheta_M)}{\partial \vartheta_i} \times \right.\right.$$
$$\left.\left. \times \overline{\frac{\partial a(t; \vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_j}}\, dt\right]_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}\right\}. \tag{6.37'}$$

En utilisant (6.36′) on peut écrire (6.35) comme suit:

$$m_1\{\zeta_i \zeta_j\} = 2 d_T^2 \operatorname{Re}\left\{\left[-\frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_j} \times \right.\right.$$
$$\times \overline{\frac{\partial \Psi(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i}} + \frac{1}{2T}\int_{-T}^{T} \frac{\partial a(t; \vartheta_1, \ldots \vartheta_M)}{\partial \vartheta_i} \times$$
$$\left.\left. \times \overline{\frac{\partial a(t; \vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_j}}\, dt\right]_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0}\right\}. \tag{6.37''}$$

En comparant (6.37′) et (6.37″) on obtient

$$m_1\{\zeta_i \zeta_j\} = -d_T^2 \left.\frac{\partial^2 h(\vartheta_1, \ldots, \vartheta_M)}{\partial \vartheta_i \partial \vartheta_j}\right|_{\vartheta_1=\vartheta_1^0, \ldots, \vartheta_M=\vartheta_M^0} \tag{6.38}$$

ou, compte tenu de (6.31),

$$m_1\{\zeta_i \zeta_j\} = -Q_{ij}. \tag{6.38'}$$

En portant (6.38') dans (6.34) on trouve

$$m_1\{(\hat{\vartheta}_k-\vartheta_k^0)(\hat{\vartheta}_n-\vartheta_n^0)\}=-\sum_{i=1}^{M}\sum_{j=1}^{M}q_{ki}q_{jn}Q_{ij}. \qquad (6.39)$$

Mais comme q_{ij} sont les éléments de la matrice $\mathbf{Q}^{-1}$ inverse de $\mathbf{Q}$, l'égalité matricielle $\mathbf{Q}^{-1}\mathbf{Q}=\mathbf{I}$ donne

$$\sum_{j=1}^{M}q_{jn}Q_{ij}=\begin{cases}1, & i=n,\\ 0, & i\neq n.\end{cases} \qquad (6.39')$$

Réunissant (6.39) et (6.39') on obtient la formule définitive donnant un élément quelconque de la matrice de corrélation des erreurs

$$m_1\{(\hat{\vartheta}_k-\hat{\vartheta}_k^0)(\hat{\vartheta}_n-\vartheta_n^0)\}=-q_{kn}. \qquad (6.40)$$

Ainsi, au signe près, les éléments de la matrice de corrélation des erreurs coïncident avec les éléments de la matrice, inverse de la matrice composée de dérivées secondes mixtes du logarithme du rapport de vraisemblance (ne différant des dérivées secondes mixtes du carré du module des fonctions d'incertitude que par le facteur d_T^2) au point correspondant aux valeurs réelles $\vartheta_1^0, \ldots, \vartheta_M^0$ des paramètres estimés. Par conséquent, la matrice de corrélation des erreurs coïncide avec la matrice, inverse de la matrice de Fisher [cf. (2.98)], c'est-à-dire que les estimations (6.30) sont *simultanément efficaces* (pour $d_T^2)\gg 1$).

6.1.4. Mesure du temps de retour d'un signal. Considérons un premier exemple illustrant la méthode du § 6.1.3 de construction des estimations asymptotiques (pour $d_T^2\to\infty$) du maximum de vraisemblance. Soit un signal à bande étroite de la forme

$$s(t;\,A,\,\varphi,\,\tau)=A\,\mathrm{Re}\,a(t-\tau)\,e^{-i\varphi}\,e^{i\omega_0 t}, \qquad (6.41)$$

où $a(t)=u(t)\,e^{i\Phi(t)}$, $u(t)$ et $\Phi(t)$ étant des fonctions réelles représentant la modulation d'amplitude et de phase du signal, et τ peut être appelé, selon la terminologie adoptée en radar, *temps de retour* du signal.

La fonction d'incertitude de ce signal est en vertu de (6.27) égale

$$\Psi(\tau-\tau_0)=\frac{1}{2T}\int_{-T}^{T}a(t-\tau_0)\,\overline{a(t-\tau)}\,dt, \qquad (6.42)$$

où τ_0 est la valeur réelle du paramètre τ. En écrivant (6.42) comme suit:

$$\Psi(\tau-\tau_0)=\frac{1}{2T}\int_{-T-\tau_0}^{T-\tau_0}a(t)\,\overline{a(t-\tau+\tau_0)}\,dt,$$

on remarque que la fonction d'incertitude coïncide, dans ce cas, avec la fonction de corrélation temporelle de l'enveloppe complexe du

signal (cf. § 4.2, tome I). La matrice Q contient maintenant un seul élément égal à [cf. (6.31) et (6.36)]

$$\frac{d^2}{d\tau^2} |\Psi(\tau-\tau_0)|^2_{\tau=\tau_0} = 2\,\mathrm{Re}\,\{\Psi''(0) - [\Psi'(0)]^2\}. \qquad (6.43)$$

Introduisons maintenant le spectre énergétique de l'enveloppe du signal [cf. (4.89), tome I]

$$F_a(\omega) = \frac{1}{T}\left|\int\limits_{-T-\tau_0}^{T-\tau_0} a(t)\,e^{-i\omega t}\,dt\right|^2. \qquad (6.44)$$

Par conséquent,

$$\Psi(\tau-\tau_0) = \frac{1}{4\pi}\int\limits_{-\infty}^{\infty} F_a(\omega)\,e^{i\omega(\tau-\tau_0)}\,d\omega, \qquad (6.45)$$

d'où

$$-2i\Psi'(0) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \omega F_a(\omega)\,d\omega = \omega^*, \qquad (6.46)$$

$$-2\Psi''(0) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \omega^2 F_a(\omega)\,d\omega = \omega_*^2, \qquad (6.47)$$

où ω^* et ω_*^2 sont la pulsation moyenne et le moment deux du spectre énergétique de l'enveloppe du signal, compte tenu de la normalisation (6.21) (les notations employées sont celles du § 8.6.1 du premier tome).

Portons (6.46) et (6.47) dans (6.43), il vient

$$\left[-\frac{d^2}{d\tau^2}|\Psi(\tau-\tau_0)|^2\right]_{\tau=\tau_0} = \omega_*^2 - \omega^{*2} = \beta^2, \qquad (6.48)$$

où β^2 est la déviation quadratique moyenne (écrat de la moyenne) de la pulsation du spectre de l'enveloppe. On prend parfois la grandeur β pour mesure quantitative de la bande du spectre de l'enveloppe.

Comme dans le cas envisagé la matrice inverse Q^{-1} ne contient qu'un seul élément égal à $-\frac{1}{d_T^2\beta^2}$, les relations (6.30), (6.30′) et (6.46) donnent l'expression asymptotique suivante (pour $d_T^2 \gg 1$) de l'estimation $\hat{\tau}$ du maximum de vraisemblance du temps de retour τ du signal :

$$\hat{\tau} - \tau_0 = \frac{2}{\beta^2}\,\mathrm{Re}\left\{e^{i\varphi_0}\frac{1}{2TA_0}\int\limits_{-T}^{T}\left[\overline{\frac{\partial a(t-\tau)}{\partial\tau}}\Bigg|_{\tau=\tau_0} - i\omega^*\overline{a(t-\tau_0)}\right]z(t)\,dt\right\}. \qquad (6.49)$$

Cette estimation est non biaisée, en vertu de (6.40) sa variance est égale à

$$M_2\{\hat{\tau}\} = \frac{1}{d_T^2\beta^2}. \qquad (6.50)$$

Pour un spectre symétrique de l'enveloppe on a $\omega^* = 0$ et l'expression (6.49) peut alors s'écrire comme suit

$$\hat{\tau} - \tau_0 = \frac{2}{\omega_*^2} \operatorname{Re}\left[e^{i\varphi_0} \frac{1}{2TA_0} \int_{-T}^{T} \overline{\frac{\partial a(t-\tau)}{\partial \tau}} \bigg|_{\tau=\tau_0} z(t)\, dt \right]. \quad (6.49')$$

A partir de (6.23′) et (6.23″) on obtient également les estimations de l'amplitude et de la phase

$$\hat{A} = \left| \frac{1}{2T} \int_{-T}^{T} a(t-\hat{\tau})\, z(t)\, dt \right|, \quad (6.51)$$

$$\hat{\varphi} = \operatorname{arc\,tg} \frac{\operatorname{Im} \int_{-T}^{T} a(t-\hat{\tau})\, z(t)\, dt}{\operatorname{Re} \int_{-T}^{T} a(t-\hat{\tau})\, z(t)\, dt}, \quad (6.51')$$

où l'on substitue $\hat{\tau}$ à partir de (6.49).

6.1.5. Mesure simultanée du temps de retour et du déplacement de fréquence dû à l'effet Doppler. Notre second exemple illustrant la méthode exposée au § 6.1.3 sera un signal à bande étroite de forme plus compliquée que (6.41), soit:

$$s(t;\ A,\ \varphi,\ \tau,\ \Omega) = A \operatorname{Re} a(t-\tau)\, e^{-i\varphi}\, e^{-i\Omega(t-\tau)}\, e^{i\omega_0(t-\tau)}. \quad (6.52)$$

Comme dans (6.41), le paramètre τ est ici le temps de retour du signal et

$$\Omega = \omega_1 - \omega_0$$

est le déplacement de fréquence dû à l'*effet Doppler*. Il y a lieu de trouver l'estimation simultanée des paramètres τ et Ω.

En vertu de (6.27) la fonction d'incertitude *) du signal (6.52) est

$$\Psi(\tau-\tau_0,\ \Omega-\Omega_0) = \frac{1}{2T} \int_{-T}^{T} a(t-\tau_0)\, \overline{a(t-\tau)}\, \times$$
$$\times\, e^{-i\Omega_0(t-\tau_0)} e^{i\Omega(t-\tau)}\, dt = e^{i\Omega(\tau-\tau_0)} \frac{1}{2T} \int_{-T-\tau_0}^{T-\tau_0} a(t)\, \overline{a(t-\tau+\tau_0)}\, e^{i(\Omega-\Omega_0)t}\, dt. \quad (6.53)$$

En introduisant le spectre de l'enveloppe du signal

$$f_T(\omega) = \int_{-T}^{T} a(t)\, e^{-i\omega t}\, dt,$$

*) La fonction $\Psi(\tau,\ \Omega)$ de deux variables (du temps et de la pulsation) a été introduite par Woodward en qualité de fonction d'incertitude [2].

on peut écrire comme suit la fonction d'incertitude:

$$\Psi(\tau-\tau_0,\ \Omega-\Omega_0)=e^{i\Omega(\tau-\tau_0)}\frac{1}{4\pi T}\int_{-\infty}^{\infty} f_T(\omega)\,\overline{f_T(\omega-\Omega+\Omega_0)}\,e^{i\omega(\tau-\tau_0)}\,d\omega. \tag{6.54}$$

Les éléments de la matrice **Q** de dimension 2×2 sont

$$Q_{11}=d_T^2\frac{\partial^2}{\partial\tau^2}\,|\Psi(\tau-\tau_0,\ \Omega-\Omega_0)|^2_{\tau=\tau_0,\ \Omega=\Omega_0}, \tag{6.55}$$

$$Q_{12}=Q_{21}=d_T^2\frac{\partial^2}{\partial\tau\,\partial\Omega}\,|\Psi(\tau-\tau_0,\ \Omega-\Omega_0)|^2_{\tau=\tau_0,\ \Omega=\Omega_0}, \tag{6.55'}$$

$$Q_{22}=d_T^2\frac{\partial^2}{\partial\Omega^2}\,|\Psi(\tau-\tau_0,\ \Omega-\Omega_0)|^2_{\tau=\tau_0,\ \Omega=\Omega_0}. \tag{6.55''}$$

En vertu de (6.54) et (6.48) on a

$$Q_{11}=d_T^2\frac{\partial^2}{\partial\tau^2}\,|\Psi(\tau-\tau_0,\ 0)|^2_{\tau=\tau_0}=-d_T^2\beta^2. \tag{6.56}$$

La grandeur Q_{22} est obtenue à partir de (6.53)

$$Q_{22}=d_T^2\frac{\partial^2}{\partial\Omega^2}\,|\Psi(0,\ \Omega-\Omega_0)|^2_{\Omega=\Omega_0}=$$

$$=d_T^2\left[\left(\frac{1}{2T}\int_{-T}^{T} t\,|a(t-\tau_0)|^2\,dt\right)^2-\frac{1}{2T}\int_{-T}^{T} t^2\,|a(t-\tau_0)|^2\,dt\right]. \tag{6.57}$$

En introduisant d'une manière analogue à ω^* et ω_*^2 le *temps moyen* t^* et le *moment deux* t_*^2 du signal et en désignant comme suit la durée quadratique moyenne du signal:

$$\alpha^2=t_*^2-(t^*)^2=\frac{1}{2T}\int_{-T}^{T} t^2\,|a(t-\tau_0)|^2\,dt-\left(\frac{1}{2T}\int_{-T}^{T} t\,|a(t-\tau_0)|^2\,dt\right)^2, \tag{6.58}$$

on peut écrire (6.57) sous la forme

$$Q_{22}=-d_T^2\,\alpha^2.$$

A partir de (6.53) et (6.54) on trouve également

$$Q_{12}=d_T^2\left\{-\frac{1}{2\pi}\int_{-\infty}^{\infty}\omega F_s(\omega)\,d\omega\cdot\frac{1}{2T}\int_{-T}^{T} t\,|a(t-\tau_0)|^2\,dt+\right.$$

$$\left.+\frac{i}{2T}\int_{-T}^{T}\left[a(t-\tau_0)\,\overline{a'(t-\tau_0)}-\overline{a(t-\tau_0)}\,a'(t-\tau_0)\right]t\,dt\right\}$$

ou

$$Q_{12}=-d_T^2\,(\omega^*t^*-\lambda^*), \tag{6.59}$$

où on a désigné par λ^* une grandeur sans dimension caractérisant la *modulation de fréquence* du signal *)

$$\lambda^* = \mathrm{Im}\,\frac{1}{T}\int_{-T}^{T} ta\,(t-\tau_0)\,\overline{a'\,(t-\tau_0)}\,dt. \qquad (6.60)$$

Ainsi, la matrice Q est de la forme

$$\mathrm{Q} = d_T^4 \begin{Vmatrix} -\beta^2 & -\omega^* t^* + \lambda^* \\ -\omega^* t^* + \lambda^* & -\alpha^2 \end{Vmatrix}. \qquad (6.61)$$

La matrice inverse est facile à calculer. Calculons le déterminant de la matrice Q

$$\det \mathrm{Q} = d_T^4\,[\alpha^2\beta^2 - (\omega^* t^* - \lambda^*)^2], \qquad (6.62)$$

qui permet de trouver la matrice de corrélation des erreurs de mesure du temps de retour du signal et du déplacement de fréquence Doppler, soit :

$$\mathrm{Q}^{-1} = \frac{1}{d_T^4\,[\alpha^2\beta^2 - (\omega^* t^* - \lambda^*)^2]^2} \begin{Vmatrix} -\alpha^2 & \omega^* t^* - \lambda^* \\ \omega^* t^* - \lambda^* & -\beta^2 \end{Vmatrix}. \qquad (6.63)$$

Les variances des estimations du temps de retour et du déplacement Doppler sont situées sur la diagonale principale de la matrice Q^{-1}, les covariances de ces estimations (au signe près) se situant sur la diagonale collatérale [cf. (6.40)].

On peut simplifier quelque peu la formule (6.63) par un choix judicieux de l'origine des coordonnées annulant dans le plan temps-pulsation les grandeurs t^* et ω^*. On a alors

$$\mathrm{Q}^{-1} = \frac{1}{d_T^4\,(\omega_*^2 t_*^2 - \lambda^{*2})^2} \begin{Vmatrix} -t_*^2 & -\lambda^* \\ -\lambda^* & -\omega_*^2 \end{Vmatrix}. \qquad (6.64)$$

En utilisant (6.30), (6.30'), (6.64) on trouve l'expression asymptotique suivante (pour $d_T^2 \gg 1$) des estimations simultanées du temps de retour et du déplacement Doppler

$$\hat{\tau} - \tau_0 = \frac{2}{\omega_*^2 t_*^2 - \lambda^{*2}}\,\mathrm{Re}\left\{ e^{i\varphi_0}\,\frac{1}{2TA_0}\,\times \right.$$

$$\left. \times \int_{-T}^{T} \left[t_*^2\, \overline{\frac{\partial a\,(t-\tau)}{\partial \tau}}\Bigg|_{\tau=\tau_0} - i\lambda^* \overline{ta\,(t-\tau_0)} \right] e^{-i\Omega_0 (t-\tau_0)} z\,(t)\,dt \right\}, \qquad (6.65)$$

$$\hat{\Omega} - \Omega_0 = \frac{2}{\omega_*^2 t_*^2 - \lambda^{*2}}\,\mathrm{Re}\left\{ e^{i\varphi_0}\,\frac{1}{2TA_0}\,\times \right.$$

$$\left. \times \int_{-T}^{T} \left[\lambda^*\, \overline{\frac{\partial a\,(t-\tau)}{\partial \tau}}\Bigg|_{\tau=\tau_0} - i\omega_*^2 \overline{ta\,(t-\tau_0)} \right] e^{-i\Omega_0 (t-\tau_0)} z\,(t)\,dt \right\}. \qquad (6.65')$$

---

*) En effet, si $a\,(t) = u\,(t)\,e^{i\Phi(t)}$, on a $\lambda^* = \dfrac{1}{T}\displaystyle\int_{-T}^{T} tu^2\,(t)\,\frac{d\Phi(t)}{dt}\,dt.$

Dans le cas général, en l'absence de modulation de fréquence du signal ($\lambda^* = 0$), les estimations du temps de retour et du déplacement Doppler deviennent non corrélées et leurs variances sont égales à [cf. (6.40) ainsi que (6.48)]

$$M_2\{\hat{\tau}\} = \frac{1}{d_T^2 \omega_*^2},$$

$$M_2\{\hat{\Omega}\} = \frac{1}{d_T^2 t_*^2}. \qquad (6.65'')$$

6.2. ESTIMATIONS DE BAYES DES PARAMÈTRES ALÉATOIRES D'UN SIGNAL

6.2.1. Estimations simultanées de l'amplitude et de la phase. Nous allons étudier les estimations des paramètres aléatoires d'un signal quasi déterministe d'après la réalisation observée du mélange additif de ce signal et d'un bruit normal stationnaire, de moyenne nulle et de fonction de corrélation $B(\tau)$.

Conformément à la théorie générale, pour une fonction de pertes quadratique (et parfois pour des conditions plus générales; cf. § 3.6.4) les estimations de Bayes d'un ensemble de paramètres interdépendants d'un processus aléatoire sont les moyennes conditionnelles de ces paramètres, lorsqu'on observe la réalisation $x(t)$ du processus sur un certain intervalle de temps [cf. (3.186)].

Au § 3.6.3 nous avons étudié en détail l'estimation de Bayes de l'amplitude aléatoire a d'un signal quasi déterministe $as(t)$ noyé dans un bruit normal additif. Nous avons montré que pour la répartition normale du paramètre l'estimation en question est la moyenne pondérée des deux grandeurs suivantes: de la moyenne a priori et de l'estimation du maximum de vraisemblance (étudiée au § 3.6.2).

Supposons maintenant que le signal soit un processus aléatoire quasi déterministe à bande étroite $s(t) = a\cos(\omega_0 t - \varphi)$ de pulsation donnée ω_0, d'amplitude a et de phase φ aléatoires dont la densité de probabilité a priori conjointe $w_2(a, \varphi)$ est donnée. Le bruit blanc est également à bande étroite. Ecrivons la réalisation $x(t)$ et le signal $s(t)$ sous la forme complexe

$$x(t) = \operatorname{Re} z(t)\, e^{i\omega_0 t}, \quad s(t;\ a,\ \varphi) = \operatorname{Re} a\, e^{-i\varphi} e^{i\omega_0 t}, \qquad (6.66)$$

la fonctionnelle de vraisemblance s'écrit alors comme suit [cf. (3.78) et (5.54)]:

$$l[z(t)\,|\,a, \varphi] = \exp\left[\operatorname{Re}\int_{-T}^{T} V(t;\ a,\ \varphi)\,\overline{z(t)}\, e^{-i\omega\, t}\, dt\right] \times$$

$$\times \exp\left[-\frac{a}{2}\operatorname{Re}\int_{-T}^{T} V(t;\ a,\ \varphi)\, e^{-i\omega_0 t} e^{i\varphi}\, dt\right], \qquad (6.67)$$

où $V(t;\ a,\ \varphi)$ est la solution de l'équation intégrale

$$\int_{-T}^{T} B_z(t-u)\, V(u\,;\, a,\, \varphi)\, du = a e^{-i\varphi} e^{i\omega_0 t}, \quad |t| \leqslant T. \tag{6.67'}$$

Introduisons la fonction $U(t) = V(t;\ a,\ \varphi)\, a e^{i\varphi}$ et écrivons (6.67) et (6.67′) comme suit:

$$l[z(t)\,|\,a,\,\varphi] = \exp\left[a \operatorname{Re} e^{-i\varphi} \int_{-T}^{T} U(t)\, \overline{z(t)}\, e^{-i\omega_0 t}\, dt\right] \times$$

$$\times \exp\left[-\frac{a^2}{2} \operatorname{Re} \int_{-T}^{T} U(t)\, e^{-i\omega_0 t}\, dt\right] =$$

$$= \exp\left[a \left| \int_{-T}^{T} U(t)\, \overline{z(t)}\, e^{-i\omega_0 t}\, dt \right| \cos(\varphi - \psi_T)\right] \times$$

$$\times \exp\left[-\frac{a^2}{2} \operatorname{Re} \int_{-T}^{T} U(t)\, e^{-i\omega_0 t}\, dt\right], \tag{6.68}$$

où

$$\psi_T = \operatorname{arc\,tg} \frac{\operatorname{Im} \int_{-T}^{T} U(t)\, \overline{z(t)}\, e^{-i\omega_0 t}\, dt}{\operatorname{Re} \int_{-T}^{T} U(t)\, \overline{z(t)}\, e^{-i\omega_0 t}\, dt}, \tag{6.68'}$$

$U(t)$ étant la solution de l'équation intégrale

$$\int_{-T}^{T} B_z(t-y)\, U(y)\, dy = e^{i\omega_0 t}, \quad |t| \leqslant T, \tag{6.69}$$

où $B_z(\tau)$ est l'enveloppe complexe de la fonction de corrélation du bruit [cf. (5.35′)].

Tout comme au § 5.2.2, introduisons les notations

$$r_T^2 = \sigma_a^2 \int_{-T}^{T} \int_{-T}^{T} U(t)\, \overline{U(y)}\, \overline{z(t)}\, z(y)\, e^{-i\omega_0 (t-y)}\, dt\, dy, \tag{6.70}$$

$$d_T^2 = \sigma_a^2 \int_{-T}^{T} \overline{U(t)}\, e^{i\omega_0 t}\, dt = \sigma_a^2 \int_{-T}^{T} \int_{-T}^{T} B_z(t-y)\, U(y)\, \overline{U(t)}\, dy\, dt > 0, \tag{6.71}$$

où σ_a^2 est le carré moyen de l'amplitude.

En utilisant (6.68) et les notations introduites on obtient les expressions suivantes pour les estimations de Bayes de l'amplitude

et de la phase d'un signal à bande étroite noyé dans un bruit additif normal à bande étroite :

$$\hat{a} = m_1\{a \mid z(t)\} = \frac{\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} a w_2(a, \varphi)\, l[z(t) \mid a, \varphi]\, d\varphi\, da}{\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} w_2(a, \varphi)\, l[z(t) \mid a, \varphi]\, d\varphi\, da} =$$

$$= \frac{\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} a w_2(a, \varphi)\, e^{-\frac{a^2 d_T^2}{2\sigma_a^2}} \exp\left[\frac{a r_T}{\sigma_a}\cos(\varphi - \psi_T)\right] d\varphi\, da}{\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} w_2(a, \varphi)\, e^{-\frac{a^2 d_T^2}{2\sigma_a^2}} \exp\left[\frac{a r_T}{\sigma_a}\cos(\varphi - \psi_T)\right] d\varphi\, da}, \qquad (6.72)$$

$$\hat{\varphi} = m_1\{\varphi \mid z(t)\} =$$

$$= \frac{\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} \varphi w_2(a, \varphi)\, e^{-\frac{a^2 d_T^2}{2\sigma_a^2}} \exp\left[\frac{a r_T}{\sigma_a}\cos(\varphi - \psi_T)\right] d\varphi\, da}{\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} w_2(a, \varphi)\, e^{-\frac{a^2 d_T^2}{2\sigma_a^2}} \exp\left[\frac{a r_T}{\sigma_a}\cos(\varphi - \psi_T)\right] d\varphi\, da}. \qquad (6.73)$$

Si le bruit additif est un bruit blanc d'intensité N_0, la solution de l'équation intégrale (6.69) est $U(t) = \frac{1}{N_0} e^{i\omega_0 t}$, les grandeurs r_T et d_T des formules (6.70) et (6.71) sont égales à

$$r_T = \frac{\sigma_a}{N_0}\left|\int_{-T}^{T} z(t)\, dt\right| = \frac{\sigma_a}{N_0}\left[\left(\int_{-T}^{T} A(t)\, dt\right)^2 + \left(\int_{-T}^{T} C(t)\, dt\right)^2\right]^{\frac{1}{2}}, \qquad (6.74)$$

où $A(t)$ et $C(t)$ sont les composantes en quadrature de la réalisation observée (cf. § 5.2.3) ; et

$$d_T^2 = \frac{2T\sigma_a^2}{N_0}. \qquad (6.74')$$

Si la phase est uniformément répartie et ne dépend pas de l'amplitude on a

$$\hat{a} = \sigma_a \frac{\int_{-\infty}^{\infty} x w_1(\sigma_a x)\, I_0(r_T x)\, e^{-\frac{d_T^2 x^2}{2}}\, dx}{\int_{-\infty}^{\infty} w_1(\sigma_a x)\, I_0(r_T x)\, e^{-\frac{d_T^2 x^2}{2}}\, dx}, \qquad (6.75)$$

$$\hat{\varphi}=\frac{\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} w_1(\sigma_a x)\, e^{-\frac{x^2 d_T^2}{2}} \left(\int\limits_{-\pi}^{\pi} \varphi \exp\left[r_T x \cos(\varphi-\psi_T)\right] d\varphi\right) dx}{\int\limits_{-\infty}^{\infty} w_1(\sigma_a x)\, I_0(r_T x)\, e^{-\frac{x^2 d_T^2}{2}}\, dx}, \qquad (6.75')$$

où $w_1(x)$ est la densité de probabilité de l'amplitude.

Pour un signal intense on peut utiliser la formule asymptotique de la fonction de Bessel et écrire comme suit l'estimation de Bayes (6.75) de l'amplitude du signal:

$$\hat{a}\sim\sigma_a\frac{\int\limits_{-\infty}^{\infty}\sqrt{x}\, w_1(\sigma_a x)\exp\left(-\frac{d_T^2 x^2}{2}+r_T x\right)dx}{\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{x}}\, w_1(\sigma_a x)\exp\left(-\frac{d_T^2 x^2}{2}+r_T x\right)dx}=$$

$$=\sigma_a\frac{\int\limits_{-\infty}^{\infty}\sqrt{x}\, w_1(\sigma_a x)\exp\left[-\frac{d_T^2}{2}\left(x-\frac{r_T}{d_T^2}\right)^2\right]dx}{\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{x}}\, w_1(\sigma_a x)\exp\left[-\frac{d_T^2}{2}\left(x-\frac{r_T}{d_T^2}\right)^2\right]dx}.$$

Pour $d_T\gg 1$ la fonction $\frac{d_T}{\sqrt{2\pi}}\exp\left[-\frac{d_T^2}{2}\left(x-\frac{r_T}{d_T^2}\right)^2\right]$ est voisine d'une fonction delta. C'est pourquoi si la densité $w_1(\sigma_a x_s)$ est continue au voisinage du point $x=\frac{r_T}{d_T^2}$ et n'est pas nulle en ce point, on obtient, compte tenu des propriétés de filtrage des fonctions delta (cf. annexe III), la formule asymptotique suivante

$$\hat{a}\sim\frac{\sigma_a r_T}{d_T^2}=\frac{\left|\int\limits_{-T}^{T} U(t)\,\overline{z(t)}\,e^{-i\omega_0 t}\,dt\right|}{\int\limits_{-T}^{T}\int\limits_{-T}^{T} B_z(t-y)\,U(y)\,\overline{U(t)}\,dy\,dt}. \qquad (6.76)$$

D'une manière analogue, on peut obtenir la formule asymptotique de l'estimation de la phase pour $d_T^2\gg 1$:

$$\hat{\varphi}\sim\psi_T=\operatorname{arc\,tg}\frac{\operatorname{Im}\int\limits_{-T}^{T}\overline{U(t)}\,z(t)\,e^{i\omega_0 t}\,dt}{\operatorname{Re}\int\limits_{-T}^{T}\overline{U(t)}\,z(t)\,e^{i\omega_0 t}\,dt}. \qquad (6.76')$$

Ainsi, les estimations de Bayes de l'amplitude et de la phase coïncident avec les estimations du maximum de vraisemblance de

ces paramètres [cf. (6.5) et (6.5′)] et donc

$$\hat{a} \sim \left| \frac{1}{2T} \int_{-T}^{T} z(t)\, dt \right|, \tag{6.76″}$$

$$\hat{\varphi} \sim \operatorname{arc\,tg} \frac{\operatorname{Im} \int_{-T}^{T} z(t)\, dt}{\operatorname{Re} \int_{-T}^{T} z(t)\, dt}, \tag{6.76‴}$$

tout comme dans les formules (6.9) et (6.7).

6.2.2. Estimations simultanées d'un nombre fini de paramètres d'un signal quasi déterministe. Considérons le problème du calcul des estimations simultanées de Bayes des paramètres d'un signal quasi déterministe sous sa forme plus générale (cf. § 3.6.6). Soit un signal de forme donnée dépendant de m paramètres aléatoires

$$s(t;\, \vartheta_1, \ldots, \vartheta_m) = \sum_{j=1}^{m} \vartheta_j s_j(t), \tag{6.77}$$

où $s_j(t)$ sont des fonctions données, et ϑ_j des paramètres aléatoires dont la densité de probabilité est égale à $w_m(\vartheta_1, \ldots, \vartheta_m)$. Sur l'intervalle $(-T, T)$ on observe la réalisation $x(t)$ de la somme de $s(t;\, \vartheta_1, \ldots, \vartheta_m)$ et d'un processus aléatoire normal de moyenne nulle et de fonction de corrélation $B(u, v)$ donnée. Il y a lieu de trouver les estimations de Bayes des paramètres $\vartheta_1, \ldots, \vartheta_m$.

Le logarithme de la fonctionnelle du rapport de vraisemblance pour $\vartheta_1, \ldots, \vartheta_m$ donnés est en vertu de (3.157) égal à

$$\ln l\,[x(t) \mid \vartheta_1, \ldots, \vartheta_m] =$$

$$= \int_{-T}^{T} V(t;\, \vartheta_1, \ldots, \vartheta_m) \left[x(t) - \frac{1}{2} s(t;\, \vartheta_1, \ldots, \vartheta_m) \right] dt, \tag{6.78}$$

où $V(t;\, \vartheta_1, \ldots, \vartheta_m)$ est la solution de l'équation intégrale

$$\int_{-T}^{T} B(t;\, u)\, V(u;\, \vartheta_1, \ldots, \vartheta_m)\, du = s(t;\, \vartheta_1, \ldots, \vartheta_m), \quad |t| \leqslant T. \tag{6.78′}$$

Portons (6.77) dans (6.78′), cette dernière équation se trouve alors remplacée par le système d'équations

$$\int_{-T}^{T} B(t, u)\, V_i(u)\, du = s_i(t), \quad |t| \leqslant T, \quad i = 1, \ldots, m, \tag{6.79}$$

avec

$$V(t;\, \vartheta_1, \ldots, \vartheta_m) = \sum_{j=1}^{m} \vartheta_j V_j(t). \tag{6.80}$$

L'expression (6.78) peut maintenant s'écrire comme suit

$$\ln l\,[x(t)\,|\,\vartheta_1, \ldots, \vartheta_m] = \sum_{j=1}^{m} \vartheta_j \int_{-T}^{T} x(t)\, V_j(t)\, dt - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \vartheta_i \vartheta_j \int_{-T}^{T} V_i(t)\, s_j(t)\, dt$$

ou, en utilisant les notations (3.206), (3.206'),

$$\ln l\,[x(t)\,|\,\vartheta_1, \ldots, \vartheta_m] = \sum_{j=1}^{m} \vartheta_j x_{Tj} - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \vartheta_i \vartheta_j s_{Tij}. \tag{6.81}$$

Compte tenu de (3.186), écrivons l'expression générale des estimations de Bayes des paramètres pour une fonction de pertes quadratique, soit:

$$\hat{\vartheta}_k = m_1\{\vartheta_k\,|\,x(t)\} = \left[\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \vartheta_k w_m(\vartheta_1, \ldots, \vartheta_m) \exp\left(\sum_{j=1}^{m} \vartheta_j x_{Tj} - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \vartheta_i \vartheta_j s_{Tij} \right) d\vartheta_1 \ldots d\vartheta_m \right] \times \left[\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w_m(\vartheta_1, \ldots, \vartheta_m) \exp\left(\sum_{j=1}^{m} \vartheta_j x_{Tj} - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \vartheta_i \vartheta_j s_{Tij} \right) d\vartheta_1 \ldots d\vartheta_m \right]^{-1},$$

$$k = 1, \ldots, m, \tag{6.82}$$

ou sous forme vectorielle

$$\hat{\boldsymbol{\vartheta}} = \frac{\int\limits_{G_{\vartheta}} \boldsymbol{\vartheta} w(\boldsymbol{\vartheta}) \exp\left(\boldsymbol{\vartheta}' \mathbf{x}_T - \frac{1}{2} \boldsymbol{\vartheta}' \mathbf{s}_T \boldsymbol{\vartheta} \right) d\boldsymbol{\vartheta}}{\int\limits_{G_{\vartheta}} w(\boldsymbol{\vartheta}) \exp\left(\boldsymbol{\vartheta}' \mathbf{x}_T - \frac{1}{2} \boldsymbol{\vartheta}' \mathbf{s}_T \boldsymbol{\vartheta} \right) d\boldsymbol{\vartheta}}. \tag{6.82'}$$

Dans le cas d'un bruit additif blanc on a $V_i(t) = \frac{1}{N_0} s_i(t)$ et

$$x_{Tj} = \frac{1}{N_0} \int_{-T}^{T} x(t)\, s_j(t)\, dt, \tag{6.83}$$

$$s_{Tij} = \frac{1}{N_0} \int_{-T}^{T} s_i(t)\, s_j(t)\, dt. \tag{6.84}$$

Si dans (6.77) les fonctions $s_j(t)$, $j = 1, \ldots, m$ sont orthogonales sur l'intervalle $(-T, T)$, on a

$$s_{Tij} = \begin{cases} \frac{E_i}{N_0}, & i=j, \\ 0, & i \neq j, \end{cases} \tag{6.85}$$

où

$$E_i = \int_{-T}^{T} s_i^2(t)\,dt \tag{6.86}$$

est l'énergie du signal $s_i(t)$ sur l'intervalle d'observation. Dans ce cas particulier, la formule (6.82) se simplifie et on a:

$$\begin{aligned} \hat{\vartheta}_k = \Big[\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} & \vartheta_k w_m(\vartheta_1, \ldots, \vartheta_m) \times \\ & \times \exp\Big[-\frac{1}{2N_0}\sum_{j=1}^{m} E_j\Big(\vartheta_j - \frac{N_0 x_{Tj}}{E_j}\Big)^2\Big] d\vartheta_1 \ldots d\vartheta_m\Big] \times \\ & \times \Big[\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w_m(\vartheta_1, \ldots, \vartheta_m) \times \\ & \times \exp\Big[-\frac{1}{2N_0}\sum_{j=1}^{m} E_j\Big(\vartheta_j - \frac{N_0 x_{Tj}}{E_j}\Big)^2\Big] d\vartheta_1 \ldots d\vartheta_m\Big]^{-1}. \end{aligned} \tag{6.87}$$

Pour $\frac{E_j}{N_0} \to \infty$, $j = 1, \ldots, m$ (signal intense),

$$\frac{E_1 \ldots E_m}{(2\pi N_0)^{m/2}} \exp\Big[-\frac{1}{2N_0}\sum_{j=1}^{m} E_j\Big(\vartheta_j - \frac{N_0 x_{Tj}}{E_j}\Big)^2\Big] \sim \prod_{j=1}^{m} \delta\Big(\vartheta_j - \frac{N_0 x_{Tj}}{E_j}\Big).$$

En utilisant la propriété de filtrage des fonctions delta, on obtient à partir de (6.87), pour $\frac{E_j}{N_0} \to \infty$, les estimations asymptotiques de Bayes des paramètres d'un signal intense

$$\hat{\vartheta}_k \sim \frac{x_{Tk}}{E_k/N_0} = \frac{\int_{-T}^{T} x(t)\,s_k(t)\,dt}{\int_{-T}^{T} s_k^2(t)\,dt}, \qquad k = 1, \ldots, m. \tag{6.88}$$

L'estimation $\hat{\vartheta}_k$ obtenue coïncide avec l'estimation du maximum de vraisemblance du paramètre ϑ_k du signal $\vartheta_k s_k(t)$ [cf. (3.168)].

Il a été montré [13] que dans le cas général aussi, on peut à partir de (6.82), pour un signal intense, obtenir la formule suivante *)

$$\hat{\boldsymbol{\vartheta}} = \mathbf{x}_T \mathbf{s}_T^{-1}, \tag{6.89}$$

c'est-à-dire que l'estimation de Bayes tend vers l'estimation du maximum de vraisemblance [cf. (3.208)]. La formule (6.88) est évidemment un cas particulier de (6.89) lorsque le bruit est blanc et les fonctions $s_j(t)$ sont orthogonales.

6.2.3. Estimation d'un signal aléatoire stationnaire noyé dans un bruit. Soit un signal à bande étroite correspondant à un processus aléatoire de moyenne nulle et de fonction de corrélation $B_s(\tau)$. Exprimons ce signal en fonction de l'enveloppe complexe (cf. § 5.2.2)

$$s(t) = \operatorname{Re} z_s(t)\, e^{i\omega_0 t} \tag{6.90}$$

et utilisons le développement orthogonal de cette enveloppe [cf. (3.36)]

$$z_s(t) = \sum_{k=1}^{\infty} \vartheta_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \qquad |t| < T, \tag{6.91}$$

où λ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale linéaire homogène

$$\varphi(t) = \lambda \int_{-T}^{T} B_{z_s}(t-y)\, \varphi(y)\, dy, \tag{6.92}$$

et $B_{z_s}(\tau)$ est l'enveloppe complexe de la fonction de corrélation du signal. Les coordonnées ϑ_k de l'enveloppe complexe du signal

$$\vartheta_k = \sqrt{\lambda_k} \int_{-T}^{T} z_s(t)\, \overline{\varphi_k(t)}\, dt, \qquad k = 1, 2, \ldots, \tag{6.92'}$$

ne sont pas corrélées. La loi de répartition des ϑ_k étant normale, la partie réelle α_k et la partie imaginaire β_k de $\vartheta_k = \alpha_k + i\beta_k$ sont également normales, de plus [cf. (5.38) à (5.38'')]

$$m_1\{\alpha_k\} = m_1\{\beta_k\} = m_1\{\alpha_k \beta_l\} = 0 \tag{6.93}$$

pour tous k et l,

$$m_1\{\alpha_k \alpha_l\} = m_1\{\beta_k \beta_l\} = 0, \quad k \neq l, \tag{6.93'}$$

$$m_1\{\alpha_k^2\} = m_1\{\beta_k^2\} = \frac{1}{2}. \tag{6.93''}$$

---

*) Les estimations de Bayes ont ces propriétés asymptotiques non seulement pour une fonction de pertes quadratique, mais également pour les fonctions de pertes de forme plus générale [14].

Le problème consiste à trouver les estimations de Bayes des coordonnées ϑ_k (ou α_k et β_k), lorsque sur l'intervalle de temps $(-T, T)$ on observe la réalisation $x(t)$ de la somme du signal $s(t)$ et d'un bruit normal stationnaire à bande étroite de fonction de corrélation $B(\tau)$.

En ne conservant que les m premiers termes du développement (6.91) et en introduisant l'enveloppe complexe de la réalisation $x(t) = \mathrm{Re}\, z(t)\, e^{i\omega_0 t}$, on peut écrire la fonctionnelle du rapport de vraisemblance pour la réalisation $z(t)$, ayant fixé $\vartheta_1, \ldots, \vartheta_m$ [cf. (5.54)]

$$l[z(t) \mid \vartheta_1, \ldots, \vartheta_m] = \exp\left[\mathrm{Re} \int_{-T}^{T} V(t; \vartheta_1, \ldots, \vartheta_m)\, \overline{z(t)}\, dt\right] \times$$
$$\times \exp\left[-\frac{1}{2}\mathrm{Re} \sum_{k=1}^{m} \frac{\vartheta_k}{\sqrt{\lambda_k}} \int_{-T}^{T} \overline{V(t; \vartheta_1, \ldots, \vartheta_m)}\, \varphi_k(t)\, dt\right], \quad (6.94)$$

où $V(t; \vartheta_1, \ldots, \vartheta_m)$ est la solution de l'équation intégrale linéaire non homogène suivante:

$$\int_{-T}^{T} B_z(t-u)\, V(u; \vartheta_1, \ldots, \vartheta_m)\, du = \sum_{k=1}^{m} \vartheta_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \quad |t| \leqslant T. \quad (6.94')$$

On peut remplacer l'équation (6.94) par le système d'équations suivant [cf. (6.79)]:

$$\int_{-T}^{T} B_z(t-u)\, V_i(u)\, du = \frac{\varphi_i(t)}{\sqrt{\lambda_i}}, \quad |t| \leqslant T, \quad (6.95)$$

avec

$$V(u; \vartheta_1, \ldots, \vartheta_m) = \sum_{j=1}^{m} \vartheta_j V_j(t). \quad (6.96)$$

Portons (6.96) dans (6.94) et écrivons comme suit la fonctionnelle du rapport de vraisemblance

$$l[z(t) \mid \vartheta_1, \ldots, \vartheta_m] = \exp\left[\mathrm{Re} \sum_{j=1}^{m} \vartheta_j \int_{-T}^{T} V_j(t)\, \overline{z(t)}\, dt\right] \times$$
$$\times \exp\left[-\frac{1}{2}\mathrm{Re} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\vartheta_i \overline{\vartheta_j}}{\sqrt{\lambda_i}} \int_{-T}^{T} \overline{V_j(t)}\, \varphi_i(t)\, dt\right]. \quad (6.97)$$

Introduisons maintenant les notations

$$x_{Tj} = \operatorname{Re} \int_{-T}^{T} V_j(t) \overline{z(t)}\, dt, \tag{6.98}$$

$$y_{Tj} = \operatorname{Im} \int_{-T}^{T} V_j(t) \overline{z(t)}\, dt, \tag{6.98'}$$

$$s_{Tij} = \operatorname{Re} \int_{-T}^{T} \overline{V_j(t)}\, \varphi_i(t)\, dt, \tag{6.99}$$

$$l_{Tij} = \operatorname{Im} \int_{-T}^{T} \overline{V_j(t)}\, \varphi_i(t)\, dt. \tag{6.99'}$$

On a alors à partir de (6.97)

$$l[z(t) \mid \alpha_1, \beta_1, \ldots, \alpha_m, \beta_m] = \exp \Big\{ \sum_{j=1}^{m} (\alpha_j x_{Tj} - \beta_j y_{Tj}) -$$

$$- \frac{1}{2} \sum_{k=1}^{m} \sum_{j=1}^{m} \frac{1}{\sqrt{\lambda_k}} [(\alpha_k \alpha_j + \beta_k \beta_j) s_{Tkj} - (\beta_k \alpha_j - \alpha_k \beta_j) l_{Tkj}] \Big\}, \tag{6.100}$$

où $\alpha_k + i\beta_k = \vartheta_k$, $k = 1, \ldots, m$.

En vertu de (6.93) à (6.93''), $2m$ coordonnées du signal $\alpha_1, \beta_1, \ldots$ $\ldots, \alpha_m.\beta_m$ forment un ensemble de $2\,m$ variables aléatoires normales indépendantes de moyennes nulles et de variances égales à 1/2. Ainsi

$$w_{2m}(\alpha_1, \beta_1, \ldots, \alpha_m, \beta_m) = \frac{1}{(\pi)^m} e^{-\sum_{k=1}^{m} (\alpha_k^2 + \beta_k^2)}. \tag{6.101}$$

En portant (6.100) et (6.101) dans (6.82) nous obtenons les estimations de Bayes des coordonnées du signal $s(t)$ [cf. (6.90)] pour une fonction de pertes quadratique

$$\hat{\alpha}_k = \frac{1}{K_m} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \alpha_k \exp \Big\{ - \sum_{j=1}^{m} (\alpha_j^2 + \beta_j^2) + \sum_{j=1}^{m} (\alpha_j x_{Tj} - \beta_j y_{Tj}) -$$

$$- \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{1}{\sqrt{\lambda_i}} [(\alpha_i \alpha_j + \beta_i \beta_j) s_{Tij} - (\beta_i \alpha_j - \alpha_i \beta_j) l_{Tij}] \Big\} \times$$

$$\times\, d\alpha_1\, d\beta_1 \ldots d\alpha_m\, d\beta_m, \tag{6.102}$$

$$\hat{\beta}_k = \frac{1}{K_m} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \beta_k \exp\Big\{ -\sum_{j=1}^{m} (\alpha_j^2 + \beta_j^2) +$$

$$+ \sum_{j=1}^{m} (\alpha_j x_{Tj} - \beta_j y_{Tj}) - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{1}{\sqrt{\lambda_i}} [(\alpha_i\alpha_j + \beta_i\beta_j) s_{Tij} -$$

$$- (\beta_i\alpha_j - \alpha_i\beta_j) l_{Tij}]\Big\} \, d\alpha_1 \, d\beta_1 \ldots d\alpha_m \, d\beta_m, \quad (6.102')$$

où

$$K_m = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Big\{ -\sum_{j=1}^{m} (\alpha_j^2 + \beta_j^2) + \sum_{j=1}^{m} (\alpha_j x_{Tj} - \beta_j y_{Tj}) -$$

$$- \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{1}{\sqrt{\lambda_i}} [(\alpha_i\alpha_j + \beta_i\beta_j) s_{Tij} - (\beta_i\alpha_j - \alpha_i\beta_j) l_{Tij}]\Big\} \times$$

$$\times \, d\alpha_1 \, d\beta_1 \ldots d\alpha_m \, d\beta_m. \quad (6.103)$$

Si le bruit additif est un bruit blanc, en vertu de (6.95) on a

$$V_i(t) = \frac{1}{N_0 \sqrt{\lambda_i}} \varphi_i(t), \quad i = 1, \ldots, m, \quad (6.104)$$

où N_0 est la densité spectrale du bruit. On obtient alors à partir de (6.98) et (6.98')

$$x_{Tj} = \frac{1}{N_0 \sqrt{\lambda_j}} \operatorname{Re} \int_{-T}^{T} \varphi_j(t) \overline{z(t)} \, dt, \quad (6.105)$$

$$y_{Tj} = \frac{1}{N_0 \sqrt{\lambda_j}} \operatorname{Im} \int_{-T}^{T} \varphi_j(t) \overline{z(t)} \, dt, \quad (6.105')$$

et à partir de (6.99), (6.99'), compte tenu du fait que les fonctions propres sur l'intervalle d'observation $(-T, T)$ sont orthogonales et normées,

$$s_{Tii} = \frac{1}{N_0 \sqrt{\lambda_i}} \int_{-T}^{T} |\varphi_i(t)|^2 \, dt = \frac{1}{N_0 \sqrt{\lambda_i}}, \quad (6.106)$$

$$s_{Tij} = 0, \; i \neq j, \; l_{Tij} = 0 \text{ pour tous } i \text{ et } j. \quad (6.106')$$

Dans ce cas les expressions (6.102) à (6.103) se simplifient considérablement et les variables d'intégration se séparent. Calculons

d'abord K_m, on a :

$$K_m = \prod_{j=1}^{m} \int_{-\infty}^{\infty} \exp\left(-\alpha_j^2 + \alpha_j x_{Tj} - \frac{\alpha_j^2}{2\lambda_j N_0}\right) d\alpha_j \times$$

$$\times \prod_{j=1}^{m} \int_{-\infty}^{\infty} \exp\left(-\beta_j^2 + \beta_j y_{Tj} - \frac{\beta_j^2}{2\lambda_j N_0}\right) d\beta_j =$$

$$= \prod_{j=1}^{m} \frac{2\pi}{2+\frac{1}{\lambda_j N_0}} \exp\left[-\frac{x_{Tj}^2 + y_{Tj}^2}{2\left(2+\frac{1}{\lambda_j N_0}\right)}\right]. \quad (6.107)$$

Il est facile de voir que dans le cas particulier d'un bruit blanc, les variables d'intégration se séparant dans (6.102), l'expression de l'estimation $\hat{\alpha}_k$ peut alors s'écrire comme suit:

$$\alpha_k = \frac{K'_{m-1}}{K_m} \exp\left[\frac{-x_{Tk}^2}{2\left(2+\frac{1}{\lambda_k N_0}\right)}\right] \int_{-\infty}^{\infty} \alpha_k \exp\left[-\frac{2+\frac{1}{\lambda_k N_0}}{2} \times\right.$$

$$\left. \times \left(\alpha_k - \frac{x_{Tk}}{2+\frac{1}{\lambda_k N_0}}\right)^2\right] d\alpha_k, \quad (6.108)$$

où K'_{m-1} est obtenu à partir de (6.107) en y omettant un facteur $\sqrt{2}\pi\left(2+\frac{1}{\lambda_k N_0}\right)^{-1/2} \exp\left[\frac{-x_{Tk}^2}{2(2+1/\lambda_k N_0)}\right]$. Enfin, en calculant l'intégrale (6.108) et après simplifications évidentes, on obtient

$$\hat{\alpha}_k = \frac{x_{Tk}}{2+\frac{1}{\lambda_k N_0}}, \quad k = 1, \ldots, m. \quad (6.109)$$

De même

$$\hat{\beta}_k = \frac{y_{Tk}}{2+\frac{1}{\lambda_k N_0}}, \quad k = 1, \ldots, m. \quad (6.109')$$

Compte tenu de (6.98), (6.98') et (6.104) on a à partir de (6.109) et (6.109')

$$\hat{\vartheta}_k = \frac{1}{2+\frac{1}{\lambda_k N_0}} \int_{-T}^{T} V_j(t)\, \overline{z(t)}\, dt = \frac{\sqrt{\lambda_k}}{1+2N_0\lambda_k} \int_{-T}^{T} \varphi_k(t)\, \overline{z(t)}\, dt \quad (6.110)$$

puis en portant (6.110) dans (6.91) on obtient l'estimation de l'enveloppe complexe du signal

$$\hat{z}_s(t) = \lim_{m\to\infty} \sum_{k=1}^{m} \frac{1}{1+2N_0\lambda_k} \varphi_k(t) \int_{-T}^{T} \varphi_k(u)\, \overline{z(u)}\, du. \quad (6.111)$$

En introduisant la fonction

$$h(t, u) = \sum_{k=1}^{\infty} \frac{\varphi_k(t)\,\varphi_k(u)}{1+2N_0\lambda_k}, \tag{6.112}$$

on peut écrire (6.111) comme suit

$$\hat{z}_s(t) = \int_{-T}^{T} h(t, u)\,\overline{z(u)}\,du. \tag{6.113}$$

ce qui permet d'obtenir l'estimation de Bayes du signal

$$\hat{s}(t) = \operatorname{Re} \hat{z}_s(t)\, e^{i\omega_0 t} = \operatorname{Re} e^{i\omega_0 t} \int_{-T}^{T} h(t, u)\,\overline{z(u)}\,du, \quad |t| \leqslant T. \tag{6.114}$$

Dans l'expression (6.114) la fonction $h(t, u)$ est déterminée par des données a priori sur la fonction de corrélation du signal et la densité spectrale du bruit, et $z(u)$ est l'enveloppe complexe de la réalisation sur l'intervalle d'observation.

6.2.4. Estimations des processus aléatoires modulant une porteuse haute fréquence noyés dans un bruit blanc additif. Supposons maintenant que le signal $s(t)$ soit une porteuse harmonique de pulsation ω_0, modulée en amplitude et en phase (ou en fréquence) par les processus aléatoires $a(t)$ et $\psi(t)$. Le problème consiste à estimer les valeurs de ces processus de modulation, d'après la réalisation de la somme du signal et d'un bruit normal additif observée sur l'intervalle $(-T, T)$. Pour ne pas encombrer l'exposé nous allons nous limiter à l'étude de deux cas particuliers où le signal est une porteuse modulée en amplitude ou en phase, bien que la méthode exposée puisse être utilisée pour des cas plus généraux.

Le cas d'un signal modulé en amplitude peut être ramené au problème étudié au § 6.2.2. En effet, posons

$$s(t) = a(t)\cos(\omega_0 t + \varphi_0). \tag{6.115}$$

En utilisant le développement orthogonal de la fonction de modulation

$$a(t) = \sum_{k=1}^{\infty} a_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}},$$

où λ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale linéaire

$$\varphi(t) = \lambda \int_{-T}^{T} B_a(t-y)\,\varphi(y)\,dy, \quad |t| \leqslant T,$$

on peut écrire $s(t)$ comme une somme analogue à (6.77), soit:

$$s(t)=\sum_{j=1}^{\infty} a_j s_j(t), \tag{6.116}$$

avec

$$s_j(t)=\frac{\varphi_j(t)}{\sqrt{\lambda_j}}\cos(\omega_0 t+\varphi_0). \tag{6.116'}$$

Le problème se réduit au calcul de l'estimation simultanée des coordonnées a_k par la méthode du § 6.2.2.

Nous allons maintenant étudier plus en détail le cas où la porteuse $A_0 \cos \omega_0 t$ est modulée en phase par le processus aléatoire normal stationnaire $\psi(t)$, de moyenne nulle, dont la fonction de corrélation $B_\psi(\tau)$ varie lentement par rapport à $\cos \omega_0\tau$, le bruit normal additif étant un bruit blanc de densité spectrale N_0 (et ne dépend pas du signal).

En utilisant le développement orthogonal de la fonction de modulation sur l'intervalle d'observation

$$\psi(t)=\sum_{j=1}^{\infty}\psi_k\frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \quad |t|\leqslant T, \tag{6.117}$$

où λ_k et $\varphi_k(t)$ sont les nombres caractéristiques et les fonctions propres de l'équation intégrale linéaire

$$\varphi(t)=\lambda\int_{-T}^{T}B_\psi(t-y)\,\varphi(y)\,dy, \quad |t|\leqslant T, \tag{6.117'}$$

on peut alors écrire comme suit le signal qui est une porteuse modulée en phase:

$$s(t)=A_0\cos\left[\omega_0 t+k_{\text{ph}}\sum_{k=1}^{\infty}\psi_k\frac{\varphi_k(t)}{\sqrt{\lambda_k}}\right], \tag{6.118}$$

où ψ_k sont des variables aléatoires normales indépendantes, de moyennes nulles et de variances unité. La constante k_{ph} donne la pente de la caractéristique linéaire du modulateur de phase. Supposant que le bruit normal additif soit blanc, on peut écrire comme suit le logarithme de la fonctionnelle du rapport de vraisemblance [cf. (6.78) et (6.78')], la réalisation $x(t)$ étant observée sur l'intervalle

$(-T, T)$:

$$\ln l\,[x(t)\,|\,\psi_1, \ldots, \psi_m, \ldots] = \frac{1}{N_0}\int_{-T}^{T} s(t) \times$$

$$\times \left[x(t) - \frac{1}{2}s(t)\right]dt = \frac{A_0}{N_0}\int_{-T}^{T} x(t) \times$$

$$\times \cos\left[\omega_0 t + k_{\text{ph}}\sum_{j=1}^{\infty}\psi_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}\right]dt - \frac{1}{2N_0}\int_{-T}^{T} s^2(t)\,dt,$$

d'où, en négligeant dans le dernier terme l'intégrale de la fonction rapidement variable $\cos[2\omega_0 t + 2\psi(t)]$ on a

$$\ln l\,[x(t)\,|\,\psi_1, \ldots, \psi_m, \ldots] = \frac{A_0}{N_0}\int_{-T}^{T} x(t) \times$$

$$\times \cos\left[\omega_0 t + k_{\text{ph}}\sum_{j=1}^{\infty}\psi_k \frac{\varphi_k(t)}{\sqrt{\lambda_k}}\right]dt - \frac{A_0^2 T}{2N_0}. \quad (6.119)$$

On peut alors écrire comme suit les estimations de Bayes des coordonnées ψ_k de la fonction de modulation pour une fonction de pertes quadratique *):

$$\hat{\psi}_k = \left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\psi_k e^{-\frac{1}{2}\sum_{j=1}^{\infty}\psi_j^2}\exp\left\{\frac{A_0}{N_0}\int_{-T}^{T} x(t)\times\right.\right.$$

$$\left.\left.\times\cos\left[\omega_0 t + k_{\text{ph}}\sum_{j=1}^{\infty}\psi_j\frac{\varphi_j(t)}{\sqrt{\lambda_j}}\right]dt\right\}d\psi_1\ldots d\psi_m\ldots\right]\times$$

$$\times\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\frac{1}{2}\sum_{j=1}^{\infty}\psi_j^2}\exp\left\{\frac{A_0}{N_0}\int_{-T}^{T} x(t)\times\right.\right.$$

$$\left.\left.\times\cos\left[\omega_0 t + k_{\text{ph}}\sum_{j=1}^{\infty}\psi_j\frac{\varphi_j(t)}{\sqrt{\lambda_j}}\right]dt\right\}d\psi_1\ldots d\psi_m\ldots\right]^{-1}. \quad (6.120)$$

L'analyse de l'expression (6.120) n'est pas possible dans le cas général. Considérons d'abord le cas où $\frac{2TA_0^2}{N_0} \ll 1$ (signal faible). Dans

*) Dans (6.120) et dans la suite, il fallait prendre un nombre fini N de variables d'intégration ψ_j et après intégration passer à la limite pour $N \to \infty$. L'écriture simplifiée utilisée n'influe pas sur le résultat définitif.

ce cas on peut négliger les termes d'ordres supérieurs à $\left(\frac{2TA_0^2}{N_0}\right)^2$. En développant l'exponentielle en série par rapport à $\frac{A_0^2T}{N_0}$ et ne conservant que les deux premiers termes on a

$$\hat{\psi}=\Big[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\psi_k e^{-\frac{1}{2}\sum_{j=1}^{\infty}\psi_j^2}\Big\{1+\frac{A_0}{N_0}\int_{-T}^{T}x(t)\times$$

$$\times\cos\Big[\omega_0t+k_{\text{ph}}\sum_{j=1}^{\infty}\psi_j\frac{\varphi_j(t)}{\sqrt{\lambda_j}}\Big]dt\Big\}\,d\psi_1\ldots d\psi_m\ldots\Big]\times$$

$$\times\Big[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum_{j=1}^{\infty}\psi_j^2}\Big\{1+\frac{A_0}{N_0}\int_{-T}^{T}x(t)\times$$

$$\times\cos\Big[\omega_0t+k_{\text{ph}}\sum_{j=1}^{\infty}\psi_j\frac{\varphi_j(t)}{\sqrt{\lambda_j}}\Big]dt\Big\}\,d\psi_1\ldots d\psi_m\Big]^{-1}.$$

Comme pour x petit

$$\frac{1}{1+x}=1-x+O(x^2)$$

et

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\psi_k e^{-\frac{1}{2}\sum_{j=1}^{\infty}\psi_j^2}\,d\psi_1\ldots d\psi_m\ldots=0,$$

on obtient à partir de (6.120) aux termes de l'ordre de $O\left(\frac{A^4T^2}{N_0^2}\right)$ près

$$\hat{\psi}_k=\frac{A_0}{N_0}\,\text{Re}\int_{-T}^{T}x(t)\,e^{i\omega_0t}\Big(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\psi_k\prod_{j=1}^{\infty}\frac{1}{\sqrt{2\pi}}\times$$

$$\times\exp\Big\{-\frac{1}{2}\Big[\psi_j^2-2k_{\text{ph}}\frac{\varphi_j(t)}{\sqrt{\lambda_j}}\,\psi_j\Big]\Big\}\,d\psi_1\ldots d\psi_m\ldots\Big)\,dt,$$

puis, en calculant l'intégrale sur les variables ψ_j on a

$$\hat{\psi}_k=\frac{A_0k_{\text{ph}}}{N_0\sqrt{\lambda_k}}\int_{-T}^{T}\varphi_k(t)\,x(t)\,e^{k_{\text{ph}}^2\sum_{j=1}^{\infty}\frac{\varphi_j^2(t)}{\lambda_j}}\cos\omega_0t\,dt. \qquad (6.121)$$

Compte tenu de [cf. (3.24′)]

$$\sum_{j=1}^{\infty}\frac{\varphi_j^2(t)}{\lambda_j}=B_\psi(0)=\sigma_\psi^2, \qquad (6.122)$$

où σ_ψ^2 est la variance du processus de modulation, on obtient à partir de (6.121)

$$\hat{\psi}_k = \frac{A_0 k_{\text{ph}}}{N_0 \sqrt{\lambda_k}} e^{k_{\text{ph}}^2 \sigma_\psi^2} \int_{-T}^{T} \varphi_k(t)\, x(t) \cos \omega_0 t\, dt. \tag{6.123}$$

En portant (6.123) dans (6.117) et utilisant (3.24), on obtient l'estimation de Bayes du processus de modulation

$$\hat{\psi}(t) = \frac{A_0 k_{\text{ph}}}{N_0} e^{k_{\text{ph}}^2 \sigma_\psi^2} \int_{-T}^{T} B_\psi(t-u)\, x(u) \cos \omega_0 u\, du, \quad |t| \leqslant T. \tag{6.124}$$

D'où

$$m_1 \{\hat{\psi}(t)\} = 0. \tag{6.125}$$

Cherchons la fonction de corrélation de l'estimation $\hat{\psi}(t)$. Comme le signal et le bruit sont indépendants, on a

$$B_x(u, v) = B_s(u, v) + N_0 \delta(v-u),$$

par conséquent,

$$B_{\hat{\psi}}(u, v) = m_1 \{\hat{\psi}(u)\, \hat{\psi}(v)\} =$$

$$= \frac{A_0^2 k_{\text{ph}}^2}{N_0^2} e^{2k_{\text{ph}}^2 \sigma_\psi^2} \int_{-T}^{T} \int_{-T}^{T} B_\psi(u-x)\, B_\psi(v-y) \times$$

$$\times [B_s(x, y) + N_0 \delta(x-y)] \cos \omega_0 x \cos \omega_0 y\, dx\, dy.$$

En négligeant comme plus haut pour $\omega \geqslant \omega_0$ les intégrales des composantes contenant $\cos \omega t$ et donc la fonction de corrélation $B_s(x, y)$ de la porteuse modulée qui est proportionnelle aux fonctions à variation rapide (cf. § 12.1.2, tome I), on obtient

$$B_{\hat{\psi}}(u, v) = \frac{A_0^2 k_{\text{ph}}^2}{N_0^2} e^{2k_{\text{ph}}^2 \sigma_\psi^2} \int_{-T}^{T} B_\psi(u-y)\, B_\psi(v-y)\, dy. \tag{6.126}$$

Pour la variance de l'estimation on a

$$\sigma_{\hat{\psi}}^2(t) = B_{\hat{\psi}}(t, t) = \frac{A_0^2 k_{\text{ph}}^2}{N_0^2} e^{2k_{\text{ph}}^2 \sigma_\psi^2} \int_{-T}^{T} B_\psi^2(u-y)\, dy \tag{6.127}$$

ou

$$\left[\frac{\sigma_{\hat{\psi}}(t)}{\sigma_\psi}\right]^2 = \frac{A_0^2 k_{\text{ph}}^2 \sigma_\psi^2}{2N_0} e^{2k_{\text{ph}}^2 \sigma_\psi^2} \int_{-T}^{T} R_\psi^2(\tau)\, d\tau, \tag{6.127'}$$

où $R_\psi(\tau)$ est le coefficient de corrélation du processus de modulation. Pour $T \to \infty$ on a

$$\left[\frac{\sigma_{\hat{\psi}}(t)}{\sigma_\psi}\right]^2 \sim \frac{A_0^2 k_{\text{ph}}^2 \sigma_\psi^2}{N_0} e^{2k_{\text{ph}}^2 \sigma_\psi^2} \int_0^\infty R_\psi^2(\tau)\, d\tau. \qquad (6.128)$$

Problèmes

6.1. Montrer que si l'amplitude d'un signal à bande étroite est répartie suivant la loi de Rayleigh

$$w_1(a) = \frac{2a}{\sigma_a^2} e^{-\frac{a^2}{\sigma_a^2}}, \quad a > 0,$$

la phase étant indépendante de l'amplitude et uniformément répartie, pour une fonction de pertes quadratique l'estimation de Bayes de l'amplitude de ce signal noyé dans un bruit normal additif est

$$\hat{a} = \sqrt{\frac{\pi}{2d_T^2 + \frac{4}{\sigma_a^2}}}\; {}_1F_1\left[-\frac{1}{2};\; 1; -\frac{r_T^2}{2\left(d_T^2 + \frac{2}{\sigma_a^2}\right)}\right], \qquad (1)$$

où ${}_1F_1$ est la fonction hypergéométrique (cf. annexe V), les grandeurs r_T^2 et d_T^2 étant données par (6.70) et (6.71).

6.2. Résoudre le problème du § 6.2.2 en supposant que les résultats des observations sont représentés non pas par la réalisation $x(t) = s(t; \vartheta_1, \ldots$ $\ldots, \vartheta_m) + \xi(t)$ mais par un échantillon discret $x = (x_1, \ldots, x_N)$ où $x_i =$ $= x(t_i)$, t_i étant les instants appartenant à l'intervalle d'observation $(-T, T)$. Montrer que les estimations de Bayes des paramètres ϑ_k peuvent s'écrire sous forme vectorielle comme suit:

$$\hat{\vartheta} = \frac{\int\limits_{G_\vartheta} \vartheta w(\vartheta) \exp\left\{-\frac{N}{2\,\mathrm{tr}\,\mathbf{K}} Q_N(\mathbf{x} - \mathbf{s}\vartheta, \mathbf{k})\right\} d\vartheta}{\int\limits_{G_\vartheta} w(\vartheta) \exp\left\{-\frac{N}{2\,\mathrm{tr}\,\mathbf{K}} Q_N(\mathbf{x} - \mathbf{s}\vartheta, \mathbf{k})\right\} d\vartheta}, \qquad (2)$$

où G_ϑ est le domaine de l'espace des paramètres où la densité de probabilité $w(\vartheta)$ est donnée; s est une matrice de dimension $N \times m$ dont les colonnes sont linéairement indépendantes (au sens algébrique). La j-ième colonne de la matrice est un vecteur de composantes $s_j(t_1), \ldots, s_j(t_N)$, $j = 1, \ldots, m$ et **k** la matrice de corrélation normée de la composante aléatoire $\xi(t)$:

$$\mathbf{k} = \frac{N\mathbf{K}}{\mathrm{tr}\,\mathbf{K}}, \qquad (2')$$

tr **K** étant la trace de la matrice de corrélation **K**,

$$Q_N(\mathbf{x} - \mathbf{s}\vartheta, \mathbf{k}) = (\mathbf{x} - \mathbf{s}\vartheta)' \mathbf{k}^{-1} (\mathbf{x} - \mathbf{s}\vartheta). \qquad (3)$$

6.3. Trouver les formules suivantes pour les éléments de la matrice d'information des estimations de l'amplitude et de la phase d'un signal à bande étroite noyé dans un bruit blanc additif:

$$I_T^{(11)}(A, \varphi) = m_1\left\{\left(\frac{\partial}{\partial A} \ln l[x(t) \mid A, \varphi]\right)^2\right\} =$$

$$= m_1\left\{\frac{16T^2}{N_0^2}\left[\left(\frac{1}{2T}\int\limits_{-T}^{T} A(t)\, dt\right) \cos\varphi + \right.\right.$$

$$+\left(\frac{1}{2T}\int_{-T}^{T}C(t)\,dt\right)\sin\varphi-A\Big]^2\Big\}=\frac{8T}{N_0}=\frac{4d_T^2}{A^2}, \quad (4)$$

$$I_T^{(22)}(A,\ \varphi)=m_1\left\{\left(\frac{\partial}{\partial\varphi}\ln l\,[x(t)\mid A,\ \varphi]\right)^2\right\}=$$

$$=m_1\left\{\frac{16T^2}{N_0^2}\Big[\left(\frac{1}{2T}\int_{-T}^{T}A(t)\,dt\right)\sin\varphi-\right.$$

$$-\left(\frac{1}{2T}\int_{-T}^{T}C(t)\,dt\right)\cos\varphi\Big]^2\Big\}=\frac{8TA^2}{N_0}=4d_T^2, \quad (5)$$

$$I_T^{(12)}(A,\ \varphi)=m_1\left\{\frac{\partial}{\partial A}\ln l\,[x(t)\mid A,\ \varphi]\,\frac{\partial}{\partial\varphi}\ln l\,[x(t)\mid A,\ \varphi]\right\}=0, \quad (6)$$

où

$$d_T^2=\frac{2TA^2}{N_0}. \quad (6')$$

Trouver à partir de (4)-(6) les éléments de la matrice inverse de la matrice d'information

$$Y_T^{(11)}(A,\ \varphi)=\frac{1}{4d_T^2}, \quad (7)$$

$$Y_T^{(22)}(A,\ \varphi)=\frac{A^2}{4d_T^2}, \quad (8)$$

$$Y_T^{(12)}(A,\ \varphi)=0, \quad (9)$$

et montrer que pour $d_T\to\infty$ les estimations du maximum de vraisemblance de l'amplitude et de la phase sont asymptotiquement efficaces.

6.4. Montrer que pour un signal à bande étroite à enveloppe gaussienne $u(t)=e^{-\alpha^2t^2}$, modulé linéairement en fréquence suivant la loi $\omega(t)=\omega_0+\lambda t$, c'est-à-dire pour un signal dont l'enveloppe complexe normée est de la forme

$$a(t)=\frac{4\alpha T}{\sqrt{2\pi}}\,e^{-(\alpha^2-i\lambda)t^2},\quad \alpha T\gg 1, \quad (10)$$

le module de la fonction d'incertitude étant égal à

$$|\Psi(\tau)|=e^{-\frac{\alpha^2}{2}m^2\tau^2}, \quad (11)$$

où m est le coefficient de compression donné par la formule

$$m=\sqrt{1+\frac{\lambda^2}{\alpha^4}} \quad (12)$$

(en l'absence de modulation de fréquence $m=1$).

En utilisant (11) montrer que la variance de l'estimation du temps de retour d'un signal, noyé dans un bruit blanc additif, lorsque le rapport d_T^2 de l'énergie du signal à la densité spectrale du bruit est grand, est égale à

$$M_2\{\hat{\tau}\}=\frac{1}{d_T^2\alpha^2m^2}. \quad (13)$$

6.5. Montrer que si l'on mesure simultanément le temps de retour et le déplacement Doppler, le module de la fonction d'incertitude bidimensionnelle

pour un signal de la forme (10) du problème 6.4 est égal à

$$|\Psi(\tau, \Omega)| = e^{-\frac{1}{2}\left(\alpha^2 m^2 \tau^2 - \frac{\lambda}{\alpha^2}\Omega\tau + \frac{\Omega^2}{4\alpha^2}\right)}. \tag{14}$$

En utilisant (14) trouver les formules suivantes pour les variances des estimations du temps de retour et du déplacement Doppler ainsi que pour le coefficient de corrélation de ces estimations, lorsque le rapport de l'énergie du signal à la densité spectrale du bruit blanc $d_T^2 \gg 1$:

$$M_2\{\hat{\tau}\} = \frac{1}{d_T^2 \alpha^2}, \tag{15}$$

$$M_2\{\hat{\Omega}\} = \frac{4\alpha^2 m^2}{d_T^2}, \tag{16}$$

$$r = \sqrt{1 - \frac{1}{m^2}}. \tag{17}$$

6.6. Supposons que l'amplitude et la phase d'un signal quasi déterministe à bande étroite [de la forme (6.66)] soient indépendantes, l'amplitude ayant une répartition a priori de Rayleigh, et la phase une répartition a priori uniforme. Démontrer que l'estimation de Bayes de l'amplitude est égale à

$$\hat{a} = \frac{\sigma_a}{\sqrt{1 + d_T^2}} \sqrt{\frac{\pi}{2}} \left[\left(1 + \frac{\rho_T^2}{2}\right) I_0\left(\frac{\rho_T^2}{4}\right) + \frac{\rho_T^2}{2} I_1\left(\frac{\rho_T^2}{4}\right)\right] e^{-\frac{\rho_T^2}{4}}, \tag{18}$$

où

$$\rho_T = \frac{r_T}{\sqrt{1 + d_T^2}}, \tag{19}$$

les grandeurs r_T^2 et d_T^2 étant données par les formules (6.70) et (6.71).

Montrer que pour $d_T^2 \to \infty$ on a

$$\hat{a} \sim \frac{\sigma_a}{d_T^2} r_T. \tag{20}$$

Vérifier que pour un bruit blanc l'estimation (20) coïncide avec l'estimation donnée par la formule (6.76).

Chapitre 7

ÉLÉMENTS DE LA THÉORIE DE CLASSIFICATION AVEC APPRENTISSAGE

7.1. PROBLÈME DE LA CLASSIFICATION AVEC APPRENTISSAGE

Le problème de la classification des résultats des observations a été formulé au début du présent volume comme le problème du choix, dans un ensemble de décisions $\gamma_0, \ldots, \gamma_m$, de la décision γ_k d'appartenance d'un ensemble de résultats d'observations $x_1, \ldots, x_n$ à l'une des $m + 1$ répartitions données $W_n(x_1, \ldots, x_n \mid s_k)$, $k = 0, \ldots, m$.

L'utilisation de tel ou tel critère pour obtenir l'algorithme optimal de classification était basée sur la connaissance des probabilités a priori d'appartenance des résultats à chacune des répartitions, la fonction de pertes étant donnée. Les répartitions a posteriori (conditionnelles) des valeurs échantillonnées, c'est-à-dire les fonctions de vraisemblance $W_n(x_1, \ldots, x_n \mid s_k)$ étaient supposées *entièrement connues*. Si les paramètres de ces fonctions étaient aléatoires, on supposait connues les densités de probabilité de ces paramètres.

Les problèmes de la détection et de la discrimination des signaux noyés dans un bruit sont des exemples de la synthèse des algorithmes optimaux de classification d'après les valeurs échantillonnées du processus observé ou une réalisation continue de ce processus. Nous avons illustré la solution de ces problèmes sur la base du critère de qualité de Bayes. La méthode de Bayes n'est pas toujours applicable à la synthèse des algorithmes optimaux de classification à cause des difficultés mathématiques qui surgissent. Même mises à part les possibilités limitées de l'appareil de programmation mathématique, on se heurte parfois à des difficultés de principe. Ceci concerne avant tout la description probabiliste de l'effet étudié que le chercheur peut ne pas connaître ou en avoir une connaissance incomplète. Ainsi, souvent on ne connaît pas les paramètres de la fonction de vraisemblance des valeurs échantillonnées ou la répartition a priori de ces paramètres, et dans certains autres cas, la forme de la fonction de vraisemblance. Le choix, éventuellement arbitraire, de la fonction de pertes impose de son côté une certaine réserve vis-à-vis de la méthode de Bayes.

Pour surmonter la difficulté de l'insuffisance de données a priori, on peut rechercher des algorithmes insensibles aux modifications des caractéristiques probabilistes des processus aléatoires étudiés.

Certaines possibilités, d'ailleurs déjà partiellement utilisées, apparaissent lors de l'étude du comportement asymptotique des algorithmes optimaux de Bayes, lorsqu'un paramètre quelconque, par exemple, le rapport signal/bruit, est petit ou augmente indéfiniment. Ces algorithmes limites sont dans certaines conditions invariables par rapport aux données a priori.

Les méthodes non paramétriques semblent ouvrir des horizons nouveaux. Certaines de ces méthodes ont déjà été mentionnées dans les premiers chapitres de ce volume. Citons les méthodes de vérification des hypothèses basées sur l'utilisation des échantillons ordonnés (cf. § 1.5.3) ou des critères de conformité (§ 2.6). Dans ces méthodes le niveau de signification reste inchangé, quelle que soit la répartition correspondant à l'hypothèse à vérifier. En particulier, dans les problèmes de détection des signaux noyés dans des bruits les méthodes non paramétriques permettent de trouver les algorithmes assurant la constance de la probabilité de fausse alarme même si la répartition des bruits varie.

La théorie des jeux offre ses moyens de surmonter les difficultés liées à l'absence de l'information a priori. Le problème de la classification est alors considéré comme un « jeu » avec un adversaire (la nature) dont la stratégie est en général inconnue. La stratégie du chercheur est basée généralement sur l'hypothèse que la nature choisit toujours la répartition des probabilités du processus aléatoire étudié la plus défavorable pour le chercheur. Cette stratégie conduit à la règle du minimax pour le choix d'une décision (cf. § 1.1.5). La théorie des décisions de Bayes peut être considérée comme un chapitre spécial de la théorie des jeux pour les cas où le « jeu » est joué contre un adversaire « faible », dont la stratégie, au sens probabiliste, est connue à l'avance du chercheur.

Ici nous n'allons pas étudier les méthodes de la statistique non paramétrique ni celles de la théorie des jeux. Ce présent chapitre est consacré à une troisième approche possible des problèmes de classification, en supposant que les répartitions des valeurs échantillonnées sont partiellement ou entièrement inconnues. Dans ce cas une période *d'apprentissage* doit précéder à la classification, c'est-à-dire la formation à la base des observations des estimations des répartitions inconnues ou des estimations des paramètres si l'on connaît l'allure de la fonction. Ces estimations sont ensuite utilisées pour la classification des résultats des observations, en remplaçant les caractéristiques probabilistes réelles inconnues des processus étudiés.

Il y a lieu de distinguer l'*apprentissage avec maître* donnant une suite d'observations étalons dont on connaît la répartition de tous les éléments (échantillon d'apprentissage classé), et l'*apprentissage sans maître* (auto-apprentissage) où les estimations mentionnées se forment d'après des échantillons d'apprentissage non classés. Parfois on limite d'avance la classe des algorithmes de classification,

dont les paramètres sont estimés d'après le critère de qualité choisi au cours de l'apprentissage ou de l'auto-apprentissage.

Dans le dispositif de réception destiné à la classification (par exemple, pour la détection ou la discrimination des signaux noyés dans un bruit) lorsque les densités de probabilité conditionnelles des résultats des observations ne sont pas données, il se produit durant l'apprentissage ou l'auto-apprentissage une modification des paramètres ou de la structure de l'algorithme de classification, optimisant ces algorithmes du point de vue du critère de qualité adopté. Ces dispositifs de réception sont dits *adaptatifs*.

Parfois, dans les problèmes de classification on utilise non pas les observations, mais certaines fonctions (ou fonctionnelles) de ces observations appelées signes informatifs de la classe (ou signes de la forme). Conformément à cette terminologie, le problème étudié se rapporte au problème général de la reconnaissance des images d'après des signes informatifs choisis. Le choix des signes informatifs dépend du problème étudié. A l'heure actuelle les méthodes générales ne sont qu'à l'état d'ébauche. Il est tout naturel que l'on désire résoudre le problème avec un nombre minimal de signes informatifs. Dans certains cas on utilise les statistiques suffisantes comme signes informatifs.

Ainsi, dans le cas le plus général le système réalisant l'algorithme de classification se compose de deux dispositifs : d'un récepteur transformant les résultats des observations en signes informatifs caractérisant l'image, et d'un dispositif de décision rapportant l'ensemble de signes formé à la sortie à l'une des classes (images).

Les valeurs échantillonnées $x_1, \ldots, x_n$ peuvent être considérées soit comme les résultats des obvervations directes, soit comme des signes informatifs.

7.2. CLASSIFICATION DANS LE CAS DES RÉPARTITIONS NORMALES

7.2.1. Moyennes inconnues. Considérons d'abord le cas très simple où après apprentissage avec maître on a obtenu un échantillon d'apprentissage classé, $x_1^{(0)}, \ldots, x_{n_0}^{(0)}$ appartenant à la classe s_0 se caractérisant par la densité de probabilité normale unidimensionnelle $w_1(x \mid s_0)$ et $x_1^{(1)}, \ldots, x_{n_1}^{(1)}$ à la classe s_1 se caractérisant par la densité de probabilité normale unidimensionnelle $w_1(x \mid s_1)$. Supposons tout d'abord que seules les moyennes de ces répartitions normales soient inconnues, les variances étant connues et égales à σ^2. Estimons les moyennes inconnues à l'aide des estimations du maximum de vraisemblance d'après l'échantillon d'apprentissage classé, c'est-à-dire

$$\hat{a}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} x_i^{(0)}, \quad \hat{a}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(1)}. \tag{7.1}$$

Nous pouvons alors, en utilisant conformément au § 1.3 ces estimations au lieu des moyennes inconnues, formuler la règle suivante de classification des observations $x_1, \ldots, x_n$, optimale vis-à-vis du critère du maximum de vraisemblance: les observations appartiennent à la classe s_1 si

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{\hat{a}_0+\hat{a}_1}{2}\right)(\hat{a}_1-\hat{a}_0) \geqslant 0 \qquad (7.2)$$

et à la classe s_0 pour l'inégalité inverse de (7.2).

Ecrivons la formule (7.2) sous la forme d'un système d'inégalités

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geqslant \frac{\hat{a}_0+\hat{a}_1}{2}, \quad \hat{a}_1 > \hat{a}_0 \qquad (7.3)$$

ou

$$\frac{1}{n}\sum_{i=1}^{n} x_i \leqslant \frac{\hat{a}_0+\hat{a}_1}{2}, \quad \hat{a}_2 < \hat{a}_0. \qquad (7.3')$$

Ainsi, l'algorithme de classification se réduit au calcul de la moyenne arithmétique des valeurs observées et à la comparaison de cette moyenne avec un seuil *dépendant des échantillons d'apprentissage.*

Le premier membre de l'expression (7.2) est le produit de variables aléatoires normales corrélées

$$y = \hat{a}_1 - \hat{a}_0, \qquad (7.4)$$

$$z = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{\hat{a}_0+\hat{a}_1}{2}. \qquad (7.4')$$

Les moyennes et les variances de ces variables sont

$$m_1\{y\} = a_1 - a_0, \quad M_2\{y\} = \sigma^2\left(\frac{1}{n_0}+\frac{1}{n_1}\right), \qquad (7.5)$$

$$m_1\{z \mid s_1\} = \frac{a_1-a_0}{2} = -m_1\{z \mid s_0\}, \qquad (7.6)$$

$$M_2\{z \mid s_1, s_0\} = \sigma^2\left(\frac{1}{n}+\frac{1}{4n_0}+\frac{1}{4n_1}\right), \qquad (7.6')$$

et le coefficient de corrélation est

$$R_{yz} = \frac{m_1\{yz\} - m_1\{y\}\, m_1\{z\}}{\sqrt{M_2\{y\}\, M_2\{z\}}} = \frac{\frac{1}{n_1}-\frac{1}{n_0}}{\sqrt{\left(\frac{1}{n}+\frac{1}{4n_0}+\frac{1}{4n_1}\right)\left(\frac{1}{n_0}+\frac{1}{n_1}\right)}}. \qquad (7.7)$$

Rappelons que a_1 et a_0 sont les moyennes *inconnues* des densités de probabilité normales $w_1(x \mid a_1)$ et $w_1(x \mid a_0)$. Notons que pour $n_0 = n_1$ les variables y et z ne sont pas corrélées.

Lorsque la taille des échantillons d'apprentissage n_0 et n_1 augmente indéfiniment, les estimations $\hat{a}_0$ et $\hat{a}_1$ convergent en probabili-

té vers les moyennes a_0 et a_1 respectivement (car la moyenne d'échantillon est une estimation consistante de la moyenne d'une répartition). De plus, compte tenu du fait que les limites en probabilité des sommes, des différences et des produits des variables aléatoires sont égales aux sommes, aux différences et aux produits des limites, on trouve à partir de (7.3) que lorsque la taille des échantillons d'apprentissage augmente indéfiniment (la taille n de l'échantillon à rapporter à l'une des deux répartitions normales mentionnées étant fixée) la règle optimale de classification tend en probabilité vers le test de Bayes de vérification d'une hypothèse simple sur la moyenne d'une variable aléatoire normale [cf. (1.67)]. Les probabilités des erreurs de classification tendent alors asymptotiquement vers les grandeurs α et β données par les formules (1.71) et (1.72).

L'algorithme de classification (7.3) peut être généralisé au cas multidimensionnel quand il s'agit de rapporter l'échantillon observé $\mathbf{X}$ à l'une des deux répartitions normales à p dimensions de moyennes vectorielles inconnues et de matrices de corrélation $\mathbf{M}_1 = \mathbf{M}_2 = \mathbf{M}$ données. Supposons qu'à la suite de l'apprentissage avec maître on ait obtenu des échantillons classés : $\mathbf{x}_1^{(0)}, \ldots, \mathbf{x}_{n_0}^{(0)}$ à partir de la première répartition et $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(1)}$ à partir de la seconde. Chaque élément de ces échantillons est un vecteur à p dimensions. Tout comme dans (7.1), estimons des moyennes vectorielles inconnues à l'aide des moyennes arithmétiques, soit

$$\hat{\mathbf{a}}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} \mathbf{x}_i^{(0)}, \quad \hat{\mathbf{a}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbf{x}_i^{(1)}. \tag{7.8}$$

L'algorithme de classification généralisant (7.2) implique la comparaison avec un seuil de la grandeur V égale à

$$V = \left(\mathbf{X} - \frac{\hat{\mathbf{a}}_0 + \hat{\mathbf{a}}_1}{2}\right)' \mathbf{M}^{-1} (\hat{\mathbf{a}}_1 - \hat{\mathbf{a}}_0). \tag{7.9}$$

La répartition de la statistique V est étudiée dans [24], [45]. Dans [32] il a été démontré en particulier, que la statistique

$$Z = \frac{4n_0n_1V}{\sqrt{(n_0+n_1)(n_0+n_1+4n_0n_1)}}$$

peut s'écrire sous la forme d'une combinaison linéaire de deux variables aléatoires indépendantes $\chi^2(p, \lambda_1)$ et $\chi^2(p, \lambda_2)$

$$Z = (1+\rho)\,\chi^2(p, \lambda_1) - (1-\rho)\,\chi^2(p, \lambda_2), \tag{7.10}$$

où $\chi^2(p, \lambda)$ est distribuée suivant une loi du χ^2 non centrée à p degrés de liberté, le paramètre de non-centralité étant λ et la grandeur ρ donnée par la formule

$$\rho = \frac{n_1 - n_0}{\sqrt{(n_0+n_1)(n_0+n_1+4n_0n_1)}}. \tag{7.10'}$$

Les paramètres de non-centralité sont donnés par les relations

$$\lambda_1 = \frac{n_0 n_1}{4(1+\rho)} \left[\frac{1}{\sqrt{n_0+n_1}} \mp \frac{1}{\sqrt{n_0+n_1+4n_0n_1}} \right]^2 d^2, \qquad (7.11)$$

$$\lambda_2 = \frac{n_0 n_1}{4(1-\rho)} \left[\frac{1}{\sqrt{n_0+n_1}} \pm \frac{1}{\sqrt{n_0+n_1+4n_0n_1}} \right]^2 d^2, \qquad (7.11')$$

où

$$d^2 = (\mathbf{a}_1 - \mathbf{a}_0)' \, \mathrm{M}^{-1} (\mathbf{a}_1 - \mathbf{a}_0). \qquad (7.12)$$

Le signe supérieur dans l'expression entre crochets correspond au cas où $m_1\{\mathrm{X}\} = \mathbf{a}_0$ et le signe inférieur au cas où $m_1\{\mathrm{X}\} = \mathbf{a}_1$.

Pour $n_0 \to \infty$ et $n_1 \to \infty$ la répartition de V tend vers la loi normale de paramètres $(\frac{1}{2} d^2, d^2)$ si $m_1\{\mathrm{X}\} = \mathbf{a}_0$ et de paramètres $\left(-\frac{1}{2} d^2, d^2\right)$ si $m_1\{\mathrm{X}\} = \mathbf{a}_1$.

7.2.2. Moyennes inconnues (apprentissage sans maître). Reprenons le problème formulé au début du § 7.2.1, avec cette seule différence que l'échantillon d'apprentissage $x_1, \ldots, x_n$ n'est pas classé. En supposant que l'apparition de chacune des deux classes s_1 et s_0 dans chaque observation est a priori équiprobable, on peut considérer chaque élément de l'échantillon d'apprentissage comme appartenant à la répartition bimodale commune suivante [27] (cf. problème 1.5):

$$w_1(x \mid a_1, a_0) = \frac{1}{2\sqrt{2\pi}\sigma} \left[e^{-\frac{(x-a_1)^2}{2\sigma^2}} + e^{-\frac{(x-a_0)^2}{2\sigma^2}} \right]. \qquad (7.13)$$

La moyenne d'une variable aléatoire répartie suivant (7.13) est

$$a = \frac{a_1 + a_0}{2}. \qquad (7.14)$$

Comme on ne connaît pas a_1 et a_0, on ne connaît pas non plus la grandeur a. Pour l'estimation de la moyenne a de la répartition (7.13) on peut prendre la moyenne échantillonnée prise pour tout l'échantillon d'apprentissage

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} x_i. \qquad (7.15)$$

En utilisant l'estimation (7.15) au lieu de la moyenne inconnue a et en adoptant pour critère le minimum de la probabilité de classification erronée (maximum de vraisemblance), on peut formuler la règle optimale suivante de classification. L'observation X appartient à la classe s_1 si

$$X \geqslant \hat{a}, \qquad (7.16)$$

et à la classe s_0 pour l'inégalité inverse de (7.16).

Lorsque n croît indéfiniment la règle formulée de classification tend asymptotiquement vers le test de Bayes de vérification de l'hypothèse simple sur la moyenne d'une variable aléatoire normale.

L'algorithme de classification (7.16) peut être généralisé au cas multidimensionnel, en conservant la symétrie sphérique de la densité de probabilité, pour le problème de l'appartenance de l'échantillon observé **X** à l'une des deux répartitions normales à p dimensions de moyennes vectorielles $\mathbf{a}_1$ et $\mathbf{a}_0$ inconnues et de matrices de corrélation $\mathbf{M}_1 = \mathbf{M}_2 = \sigma^2\mathbf{I}$ où **I** est la matrice unité. Dans ce cas la densité de probabilité multidimensionnelle commune des deux classes est

$$w_p(\mathbf{X} \mid \mathbf{a}_1, \mathbf{a}_0) = \frac{1}{2(2\pi)^{\frac{p}{2}}\sigma^p}\left\{\exp\left[-\frac{1}{2\sigma^2}(\mathbf{X}-\mathbf{a}_1)'(\mathbf{X}-\mathbf{a}_1)\right]+\right.$$

$$\left.+\exp\left[-\frac{1}{2\sigma^2}(\mathbf{X}-\mathbf{a}_0)'(\mathbf{X}-\mathbf{a}_0)\right]\right\} =$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}\sigma^p}\exp\left[-\frac{1}{2\sigma^2}\mathbf{b}'\mathbf{b}\right]\exp\left[-\frac{1}{2\sigma^2}(\mathbf{X}-\mathbf{a})'(\mathbf{X}-\mathbf{a})\right]\times$$

$$\times \operatorname{ch}\frac{\mathbf{b}'(\mathbf{X}-\mathbf{a})}{\sigma^2}, \qquad (7.17)$$

où

$$\mathbf{a} = \frac{\mathbf{a}_1+\mathbf{a}_0}{2}; \quad \mathbf{b} = \frac{\mathbf{a}_1-\mathbf{a}_0}{2}. \qquad (7.17')$$

Le vecteur **a** est le vecteur des moyennes de la répartition (7.17), les éléments de la matrice de corrélation **M** de cette répartition étant

$$M_{ij} = \int_{-\infty}^{\infty}\cdots\int (x_i-a_i)(x_j-a_j)\,w_p(\mathbf{X} \mid \mathbf{a}, \mathbf{b})\,dx_1\ldots dx_p =$$

$$= b_ib_j + \sigma^2\delta_{ij}, \qquad (7.18)$$

où $\delta_{ij} = 1$ pour $i = j$ et $\delta_{ij} = 0$ pour $i \neq j$,
a_i, $i = 1, \ldots, p$ sont les composantes du vecteur **a**,
b_j, $j = 1, \ldots, p$ sont les composantes du vecteur **b**.

Quand les moyennes vectorielles $\mathbf{a}_1$ et $\mathbf{a}_0$ sont connues, le test optimal de Bayes de partition de l'espace des échantillons correspond à un hyperplan perpendiculaire à la ligne réunissant les points $\mathbf{X} = \mathbf{a}_1$ et $\mathbf{X} = \mathbf{a}_0$ et divisant cette dernière en deux (cf. problème 1.6). L'observation **X** appartient à l'une ou l'autre classe suivant le signe de la grandeur $\mathbf{b}'(\mathbf{X} - \mathbf{a})$ [comparer avec (7.9)].

Dans le cas où les moyennes vectorielles **a** et **b** des deux classes sont inconnues, il faut les remplacer par les estimations obtenues lors de l'apprentissage. En cas d'auto-apprentissage d'après un échantillon $\mathbf{x}_1, \ldots, \mathbf{x}_n$ non classé, ces estimations sont obtenues à partir de la moyenne échantillonnée et de la matrice de corrélation de l'échantillon. L'estimation de la moyenne vectorielle de la répartition (7.17) est égale à

$$\hat{\mathbf{a}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \qquad (7.19)$$

et on peut trouver les estimations du vecteur **b** à partir d'un système d'équations obtenu par comparaison des grandeurs M_{ij} de (7.18) avec les éléments correspondants de la matrice de corrélation de l'échantillon

$$\hat{\mathbf{M}}=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\hat{\mathbf{a}})(\mathbf{x}_i-\hat{\mathbf{a}})'. \tag{7.20}$$

Notons que le nombre d'équations de ce système sera en général supérieur au nombre d'inconnues. Pour la solution de ce système on peut utiliser la méthode des moindres carrés.

Le calcul de l'estimation du vecteur **b** peut se simplifier si l'on tient compte du fait que ce vecteur est le plus grand des vecteurs propres de la matrice de corrélation de la répartition (7.17), de plus, pour le cas envisagé de symétrie sphérique, tous les nombres caractéristiques de la matrice de corrélation, à l'exception du plus grand, sont égaux entre eux.

Le problème devient bien beaucoup plus simple lorsque la moyenne vectorielle pour l'une des classes, par exemple $\mathbf{a}_0$, est donnée (comme dans le problème de la détection d'un signal inconnu noyé dans un bruit pour $\mathbf{a}_0\equiv 0$). On a alors

$$\hat{\mathbf{b}}=\hat{\mathbf{a}}-\mathbf{a}_0=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i-\mathbf{a}_0. \tag{7.21}$$

7.2.3. Moyennes et matrices de corrélation inconnues. Supposons d'abord que les moyennes vectorielles des deux répartitions normales à p dimensions soient données et égales entre elles $\mathbf{a}_0=\mathbf{a}_1=\mathbf{a}$, et les matrices de corrélation de ces répartitions $\mathbf{M}_0$ et $\mathbf{M}_1$ soient inconnues ($\mathbf{M}_0\neq\mathbf{M}_1$). Ayant à notre disposition des échantillons d'apprentissage classés $\mathbf{x}_1^{(0)},\ldots,\mathbf{x}_{n_0}^{(1)}$ de la première répartition et $\mathbf{x}_1^{(1)},\ldots$ $\ldots,\mathbf{x}_{n_1}^{(1)}$ de la seconde, on peut écrire les estimations du maximum de vraisemblance des matrices inconnues $\mathbf{M}_0$ et $\mathbf{M}_1$ [cf. (2.203)]

$$\hat{\mathbf{M}}_0=\frac{1}{n_0}\sum_{i=1}^{n_0}(\mathbf{x}_i^{(0)}-\mathbf{a})(\mathbf{x}_i^{(0)}-\mathbf{a})', \tag{7.22}$$

$$\hat{\mathbf{M}}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}(\mathbf{x}_i^{(1)}-\mathbf{a})(\mathbf{x}_i^{(1)}-\mathbf{a})'. \tag{7.22'}$$

Pour classer l'observation **X** on peut utiliser l'algorithme optimal de vérification des hypothèses concernant la matrice de corrélation de la répartition normale (cf. problème 1.7) en remplaçant les matrices de corrélation inconnues $\mathbf{M}_0$ et $\mathbf{M}_1$ par leurs estimations (7.22) et (7.22′). On obtient alors la règle suivante de classification : l'ob-

servation **X** appartient à la seconde répartition si on a

$$(\mathbf{X}-\mathbf{a})'(\hat{\mathbf{M}}_1-\hat{\mathbf{M}}_0)(\mathbf{X}-\mathbf{a})\geqslant 2\ln c+\ln\frac{\det\hat{\mathbf{M}}_0}{\det\hat{\mathbf{M}}_1}. \qquad (7.23)$$

En introduisant le changement de variable

$$\mathbf{Y}=\hat{\mathbf{f}}'(\mathbf{X}-\mathbf{a}),\quad \mathbf{Y}'=(y_1,\ \ldots,\ y_p),$$

où la matrice $\hat{\mathbf{f}}$ est donnée par la relation

$$\hat{\mathbf{M}}_1\hat{\mathbf{f}}=\hat{\mathbf{M}}_0\hat{\mathbf{f}}\hat{\mathbf{\Lambda}}$$

et $\hat{\mathbf{\Lambda}}$ est une matrice diagonale dont les éléments $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ sont les racines de l'équation

$$\det[\hat{\mathbf{M}}_1-\hat{\lambda}\hat{\mathbf{M}}_0]=0,$$

on peut écrire comme suit l'inégalité (7.23):

$$\sum_{i=1}^{p}\left(1-\frac{1}{\hat{\lambda}_i}\right)y_i^2\geqslant 2\ln c+\sum_{i=1}^{p}\ln\hat{\lambda}_i. \qquad (7.24)$$

La répartition de la statistique dans le premier membre de (7.24), ainsi que les probabilités d'erreurs, correspondant à la règle (7.24), sont étudiées dans [39]. En particulier, pour $n_0\to\infty$ et $n_1\to\infty$ la règle (7.24) tend en probabilité vers le test optimal de vérification des hypothèses pour $\mathbf{M}_0$ et $\mathbf{M}_1$ données (cf. problème 1.7).

Quand les moyennes vectorielles des deux répartitions normales sont égales entre elles et ne sont pas données, dans (7.23) doit figurer l'estimation de **a** faite d'après les échantillons d'apprentissage

$$\hat{\mathbf{a}}=\frac{n_0\hat{\mathbf{a}}_0+n_1\hat{\mathbf{a}}_1}{n_0+n_1}, \qquad (7.25)$$

où les estimations $\hat{\mathbf{a}}_0$ et $\hat{\mathbf{a}}_1$ sont données par (7.8).

Lorsque les moyennes vectorielles des deux répartitions normales ne sont pas égales entre elles ($\mathbf{a}_0\neq\mathbf{a}_1$) et sont inconnues, et que les matrices de corrélation inconnues sont égales entre elles ($\mathbf{M}_0=\mathbf{M}_1=\mathbf{M}$) on utilise la statistique de classification (7.9), en portant dans son expression au lieu de **M** son estimation faite d'après les échantillons d'apprentissage, c'est-à dire [cf. (2.207)]

$$V=\left(\mathbf{X}-\frac{\hat{\mathbf{a}}_0+\hat{\mathbf{a}}_1}{2}\right)'\hat{\mathbf{M}}^{-1}(\hat{\mathbf{a}}_1-\hat{\mathbf{a}}_0), \qquad (7.26)$$

où

$$\hat{\mathbf{M}}=\frac{1}{n_0+n_1-2}\times$$
$$\times\left[\sum_{i=1}^{n_0}(\mathbf{x}_i^{(0)}-\hat{\mathbf{a}}_0)(\mathbf{x}_i^{(0)}-\hat{\mathbf{a}}_0)'+\sum_{i=1}^{n_1}(\mathbf{x}_i^{(1)}-\hat{\mathbf{a}}_1)(\mathbf{x}_i^{(1)}-\hat{\mathbf{a}}_1)'\right] \qquad (7.27)$$

et les estimations $\hat{\mathbf{a}}_0$, $\hat{\mathbf{a}}_1$ sont données par (7.8). La répartition de la statistique (7.26) est étudiée dans [24], [34], [45].

7.2.4. Nombre arbitraire de répartitions. Les raisonnements suivants permettent d'utiliser l'apprentissage avec maître dans le cas de $m+1$ classes $s_0, \ldots, s_m$ caractérisées par des densités de probabilité à p dimensions (qui ne sont pas obligatoirement normales) $w_p(\mathbf{x} \mid s_0), \ldots, w_p(\mathbf{x} \mid s_m)$. On connaît la forme de chacune des répartitions, dont les paramètres, tous ou en partie, sont inconnus. Si l'on dispose d'un échantillon d'apprentissage classé, c'est-à-dire de $m+1$ ensembles de vecteurs à p dimensions et l'on sait que $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n_k}^{(k)}$ appartient à la classe s_k, $k=0, \ldots, m$, on peut d'après les échantillons d'apprentissage trouver les estimations des paramètres inconnus des répartitions et en substituant ces estimations dans les expressions $w_p(\mathbf{x} \mid s_k)$ trouver les estimations $\hat{w}_p(\mathbf{x} \mid s_k)$ des fonctions des densités de probabilité (cf. § 2.6.6). Puis, en utilisant le minimum de risque moyen en tant que critère de qualité de la classification (pour des probabilités a priori p_k, $k=0, \ldots, m$ d'appartenance de l'échantillon observé à des répartitions en question et des valeurs de pertes Π_{ij} données) on peut, en remplaçant dans le test de Bayes (1.144) les densités de probabilité $w_p(\mathbf{x} \mid s_k)$ par leurs estimations, obtenir la règle suivante de classification : l'échantillon observé $\mathbf{X}$ appartient à la classe s_k si

$$\sum_{i=0}^{m} (\Pi_{ij}-\Pi_{ik})\, p_i \hat{w}_p(\mathbf{X} \mid s_i) \geqslant 0, \quad j=0, \ldots, m\,;\ j \neq k. \tag{7.28}$$

Supposons, par exemple, que l'on ait $m+1$ classes caractérisées par des densités de probabilité normales multidimensionnelles, de moyennes inconnues $\mathbf{a}_k$ et de matrices de corrélation $\mathbf{M}_k = \mathbf{M}$ $(k=0, \ldots, m)$ et que l'apprentissage fournisse des échantillons de chacune des répartitions. Si l'échantillon $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n_k}^{(k)}$ suit une loi normale de paramètres $\mathbf{a}_k$, $\mathbf{M}$ $(k=0, \ldots, m)$, on peut trouver les estimations du maximum de vraisemblance des paramètres inconnus, soit :

$$\hat{\mathbf{a}}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^{(k)}, \tag{7.29}$$

$$\hat{\mathbf{M}} = \frac{1}{\sum\limits_{i=0}^{m} n_i - (m+1)} \sum_{k=0}^{m} \sum_{i=1}^{n_k} (\mathbf{x}_i^{(k)} - \hat{\mathbf{a}}_k)(\mathbf{x}_i^{(k)} - \hat{\mathbf{a}}_k)', \tag{7.30}$$

et trouver les estimations des densités de probabilité mentionnées [cf. (2.201)] :

$$\hat{w}_p(\mathbf{x} \mid s_k) = \frac{1}{(2\pi)^{p/2} \sqrt{\det \hat{\mathbf{M}}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\hat{\mathbf{a}}_k)' \hat{\mathbf{M}}^{-1} (\mathbf{x}-\hat{\mathbf{a}}_k)\right], \tag{7.31}$$

$$k=0, \ldots, m.$$

En posant $\Pi_{ij} = \Pi \; (i \neq j)$, $\Pi_{jj} = 0$ on peut, à partir de (7.28), trouver la règle suivante de classification : l'échantillon observé **X** appartient à la classe s_k si

$$V_{kj} = \ln \frac{\hat{w}_p(\mathbf{x} \mid s_k)}{\hat{w}_p(\mathbf{x} \mid s_j)} = \left[\mathbf{x} - \frac{1}{2}(\hat{\mathbf{a}}_k + \hat{\mathbf{a}}_j) \right]' \hat{\mathbf{M}}^{-1} (\hat{\mathbf{a}}_k - \hat{\mathbf{a}}_j) \geqslant \ln \frac{p_j}{p_k},$$

$$j = 0, \ldots, m; \; j \neq k. \tag{7.32}$$

Lorsque tous les $n_k \to \infty$, la distribution conjointe des statistiques V_{kj} devient normale (voir [1], page 208).

Lorsque la forme de la fonction $w_p(\mathbf{x} \mid s_k)$ est inconnue, on peut utiliser les échantillons d'apprentissage pour estimer ces fonctions d'après la méthode donnée au § 2.7.4. En vertu de (2.211)

$$\hat{w}_p(\mathbf{x} \mid s_k) = \frac{1}{n_k} \sum_{i=1}^{n_k} \prod_{j=1}^{p} \frac{1}{h(n)} K\left[\frac{x_j - x_{ij}^{(k)}}{h(n)} \right], \tag{7.33}$$

$$\mathbf{x} = (x_1 \ldots, x_p), \; \mathbf{x}_i^{(k)} = (x_{i1}^{(k)}, \ldots, x_{ip}^{(k)}), \tag{7.34}$$

le noyau $K(y)$ de l'approximation et la grandeur $h(n)$ satisfaisant aux conditions (2.211′).

7.2.5. Relation avec la méthode géométrique. L'hypothèse selon laquelle les classes sont caractérisées par des répartitions normales reste intéressante, même si en réalité les répartitions ne sont pas normales. Les règles de classification étudiées ci-dessus, basées sur des répartitions normales de paramètres inconnus, restent optimales dans le cas des répartitions qui ne sont pas normales, si l'on prend pour critère de qualité le minimum de « distance » entre le vecteur des échantillons observés et le vecteur des échantillons d'apprentissage [4]. On appelle distance du vecteur **X** jusqu'au vecteur de l'échantillon d'apprentissage $\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_{n_i}^{(i)}$ la grandeur

$$\frac{1}{n_i} \sum_{k=1}^{n_i} \| \mathbf{X} - \mathbf{x}_k^{(i)} \|^2,$$

(où $\| \mathbf{X} - \mathbf{x}_k^{(i)} \|$ est la norme du vecteur $\mathbf{X} - \mathbf{x}_k^{(i)}$), c'est-à-dire la somme des carrés des composantes de ce vecteur.

Avant la classification, il est rationnel d'appliquer une transformation linéaire afin de comprimer au maximum les échantillons d'apprentissage d'une classe donnée. Plus exactement, il faut trouver la transformation linéaire $\mathbf{A}_i$ conservant le volume (c'est-à-dire telle que son jacobien soit égal à l'unité), rendant minimale la grandeur

$$r_i = \frac{1}{n_i(n_i - 1)} \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} \| \mathbf{A}_i (\mathbf{x}_k^{(i)} - \mathbf{x}_j^{(i)} \|^2. \tag{7.35}$$

Il a été montré dans [4] que la transformation cherchée est une rotation $\mathbf{C}_i'$ avec transformation diagonale ultérieure $\mathbf{D}_i$. Les colonnes

de la matrice $\mathbf{C}_i$ sont les vecteurs propres de la matrice de corrélation de l'échantillon $\hat{\mathbf{M}}_i$ [cf. (7.22)], les éléments de la matrice diagonale $\mathbf{D}_i$ étant

$$d_{kj}^{(i)} = \frac{1}{\sigma_k^{(i)}} \left(\prod_{l=1}^{p} \sigma_l^{(i)} \right)^{\frac{1}{p}} \delta_{kj}, \tag{7.36}$$

où $\sigma_l^{(i)}$ est l'écart quadratique moyen des vecteurs $\mathbf{x}_k^{(i)}$ dans la direction du l-ième vecteur propre de la matrice $\mathbf{C}_i$; p la dimension des vecteurs; $\delta_{kj} = 1$ pour $k = j$; $\delta_{kj} = 0$ pour $k \neq j$.

On peut maintenant formuler la règle de classification de la manière suivante: d'après les échantillons d'apprentissage et l'échantillon $\mathbf{X}$ on calcule les grandeurs

$$\rho_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \| \mathbf{D}_i \mathbf{C}_i' (\mathbf{X} - \mathbf{x}_k^{(i)}) \|^2, \quad i = 0, 1, \ldots, m, \tag{7.37}$$

et l'on choisit le plus petit d'entre eux $\rho_j = \min\limits_i \rho_i$; l'échantillon est alors rapporté à la classe j.

Comme démontré dans [4], la règle mentionnée est équivalente à celle qui est basée sur le calcul du logarithme du rapport de vraisemblance en supposant les répartitions normales, de moyennes vectorielles échantillonnées et de matrices de corrélation de l'échantillon (avec probabilités a priori $p_i = p = \frac{1}{m+1}$ et coûts $\Pi_{ij} = \Pi$).

7.3. PROCÉDURES D'APPRENTISSAGE DE BAYES

7.3.1. Méthode des probabilités a posteriori. Une autre méthode de solution du problème de la classification avec maître suppose que les paramètres $\theta_1, \ldots, \theta_n$ donnant les densités de probabilité $w_p(\mathbf{x} \mid \theta_1; s_1), \ldots, w_p(\mathbf{x} \mid \theta_n; s_m)$ des classes sont des vecteurs aléatoires indépendants, de dimension finie, de densités de probabilité a priori données $w_M(\theta_k)$, $k = 1, \ldots, m$. A partir d'un ensemble d'échantillons d'apprentissage $\mathbf{X}_{\text{appr}} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ où les éléments $\mathbf{x}_k$ sont des vecteurs lignes $\mathbf{x}_k = (\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n_k}^{(k)})$ des échantillons classés appartenant à la classe s_k et de l'observation $\mathbf{X}$, on peut, en utilisant la formule de Bayes, trouver la probabilité a posteriori de la classe s_k pour $\mathbf{X}_{\text{appr}}$, $\mathbf{X}$ donnés:

$$P\{s_k \mid \mathbf{X}_{\text{appr}}, \mathbf{X}\} = \frac{p_k W(\mathbf{X}_{\text{appr}}, \mathbf{X} \mid s_k)}{\sum\limits_{k=1}^{m} p_k W(\mathbf{X}_{\text{appr}}, \mathbf{X} \mid s_k)}, \tag{7.38}$$

où p_k est la probabilité a priori d'appartenance à la classe s_k. Connaissant les grandeurs $P\{s_k \mid \mathbf{X}_{\text{appr}}, \mathbf{X}\}$, $k = 1, \ldots, m$ on rap-

porte l'observation $\mathbf{X}$ à la classe s_j pour laquelle la probabilité a posteriori est maximale, c'est-à-dire

$$P\{s_j \mid \mathbf{X}_{appr}, \mathbf{X}\} > P\{s_k \mid \mathbf{X}_{appr}, \mathbf{X}\} \tag{7.39}$$

pour tous les $k \neq j$.

Comme

$$W(\mathbf{X}_{appr}, \mathbf{X} \mid s_k) = W(\mathbf{X} \mid \mathbf{X}_{appr}, s_k)\, W(\mathbf{X}_{appr} \mid s_k), \tag{7.40}$$

le second facteur dans (7.40) ne dépendant que des échantillons d'apprentissage, il découle de (7.38) et (7.39) que l'algorithme de la classification se ramène au calcul des grandeurs $p_k W(\mathbf{X} \mid \mathbf{X}_{appr}, s_k)$, $k = 1, \ldots, m$ et au rangement de $\mathbf{X}$ dans la classe s_j à laquelle correspond la valeur maximale obtenue. Si les probabilités a posteriori p_k sont les mêmes, la classification pour un ensemble donné $\mathbf{X}_{appr}$ revient à trouver la classe s_j pour laquelle l'échantillon observé $\mathbf{X}$ maximise la fonction de vraisemblance $W(\mathbf{X} \mid \mathbf{X}_{appr}, s_k)$. Cette dernière peut servir d'estimation des répartitions inconnues des classes pour un ensemble donné d'échantillons d'apprentissage $\mathbf{X}_{appr}$.

Calculons la fonction $W(\mathbf{X} \mid \mathbf{X}_{appr}, s_k)$ en utilisant la formule de la probabilité totale, soit:

$$W(\mathbf{X} \mid \mathbf{X}_{appr}, s_k) = \int_{\Omega_k} W(\mathbf{X} \mid \mathbf{X}_{appr}, \theta_k, s_k)\, W(\theta_k \mid \mathbf{X}_{appr}, s_k)\, d\theta_k =$$

$$= \int_{\Omega_k} W(\mathbf{X} \mid \theta_k, s_k)\, W(\theta_k \mid \mathbf{x}_k, s_k)\, d\theta_k, \quad \theta_k \in \Omega_k, \tag{7.41}$$

en effet $W(\theta_k \mid \mathbf{X}_{appr}, s_k) = W(\theta_k \mid \mathbf{x}_k, s_k)$ et $W(\mathbf{X} \mid \mathbf{X}_{appr}, \theta_k, s_k)$ ne dépend pas des échantillons d'apprentissage.

La densité de probabilité a posteriori du paramètre θ_k pour un échantillon d'apprentissage donné peut être calculée à partir de la formule de Bayes, soit:

$$W(\theta_k \mid \mathbf{x}_k, s_k) = w(\theta_k) \frac{W(\mathbf{x}_k \mid \theta_k, s_k)}{\int_{\Omega_k} w(\theta_k)\, W(\mathbf{x}_k \mid \theta_k, s_k)\, d\theta_k}. \tag{7.42}$$

Dans la formule (7.42) le second facteur indique comment varie la densité de probabilité primitive a priori du paramètre θ_k au cours de l'apprentissage avec maître.

Notons que dans le cas où les paramètres θ_k sont donnés et égaux à θ_k^*, leurs densités de probabilités conditionnelles sont des fonctions delta

$$W(\theta_k \mid \mathbf{x}_k, s_k) = \delta(\theta_k - \theta_k^*)$$

et en vertu de (7.41) on a

$$W(\mathbf{X} \mid \mathbf{X}_{appr}, s_k) = W(\mathbf{X} \mid \theta_k^*, s_k),$$

comme il fallait s'attendre. Dans ce cas le problème de la classification devient celui de la vérification des hypothèses à alternatives multiples que nous avons passé en revue dans le premier chapitre.

7.3.2. Détection d'un signal inconnu noyé dans un bruit normal. A titre d'exemple simple *) illustrant la méthode mentionnée ci-dessus nous allons étudier le problème de la détection d'un signal inconnu noyé dans un bruit additif normal non corrélé, de moyenne nulle et de variance σ^2. Le signal est une variable aléatoire normale a, indépendante du bruit, dont la répartition a priori est donnée par les paramètres (a_0, σ_0^2). Le problème de la détection d'un signal consiste à rapporter l'observation X soit au mélange du signal et du bruit $[w_1(x \mid s_1)]$ soit au bruit $[w_1(x \mid s_0)]$.

Si la densité de probabilité a priori du signal ne change pas durant l'apprentissage, c'est-à-dire si

$$W(a \mid x_1, \ldots, x_n; s_1) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{(a-a_0)^2}{2\sigma_0^2}},$$

en vertu de (7.41) les fonctions de vraisemblance $W_1(X \mid s_1)$, $W_1(X \mid s_0)$ sont **)

$$W_1(X \mid s_1) = \frac{1}{2\pi\sigma_0\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(a-a_0)^2}{2\sigma_0^2}\right] \exp\left[-\frac{(X-a)^2}{2\sigma^2}\right] da =$$

$$= \frac{1}{\sqrt{2\pi(\sigma_0^2+\sigma^2)}} \exp\left[-\frac{(X-a_0)^2}{2(\sigma_0^2+\sigma^2)}\right], \tag{7.43}$$

$$W_1(X \mid s_0) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{X^2}{2\sigma^2}\right]. \tag{7.43'}$$

Le test (règle de classification) basé sur le critère du maximum de la probabilité a posteriori est alors un cas particulier des tests de Bayes de vérification des hypothèses (de détection d'un signal) étudiés dans les chapitres 1 et 5. On prend la décision sur la présence du signal si

$$W_1(X \mid s_1) \geqslant W_1(X \mid s_0), \tag{7.44}$$

dans le cas contraire on décide que l'on observe un bruit pur. En vertu de (7.43), (7.43') et (7.44) l'algorithme de classification peut être écrit comme suit : l'observation X doit être rapportée au mélange du signal et du bruit si

$$a_0 X + \frac{1}{2}\frac{\sigma_0^2}{\sigma^2} X^2 \geqslant c, \tag{7.45}$$

---

*) Le problème général de la discrimination des signaux inconnus ($m \geqslant 2$) noyés dans des bruits est étudié dans [2].

**) La formule (7.43) découle directement du fait que la somme du signal et du bruit est dans ce cas répartie normalement, de moyenne a_0 et de variance $\sigma_0^2 + \sigma^2$.

et doit être rapportée au bruit si

$$a_0 X + \frac{1}{2}\frac{\sigma_0^2}{\sigma^2} X^2 < c, \tag{7.45'}$$

où

$$c = a_0^2 + \frac{\sigma_0^2 + \sigma^2}{2} \ln\left(1 + \frac{\sigma_0^2}{\sigma^2}\right). \tag{7.45''}$$

S'il y a lieu de classer non pas une seule observation mais un échantillon $(X_1, \ldots, X_N)$, en supposant toujours le signal et le bruit indépendants, il faut remplacer dans (7.45) et (7.45') X et X^2 par les sommes des valeurs observées et des carrés de ces valeurs [cf. (1.127) et (5.182)]. Le dispositif optimal de détection d'un signal se compose d'un filtre adapté et d'un récepteur énergétique [cf. (5.87)]. Les valeurs de sortie sont sommées, la somme obtenue étant ensuite comparée avec un seuil.

Considérons maintenant le cas où les données a priori sur le signal sont *précisées pendant le processus d'apprentissage avec maître*. Soit l'échantillon $x_1, \ldots, x_n$ appartenant au mélange d'un signal et d'un bruit. La densité de probabilité a posteriori du signal est donnée par la formule (2.168)

$$W_1(a \mid x_1, \ldots, x_n;\ s_1) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)^{1/2} \times$$
$$\times \exp\left\{-\frac{n}{2\sigma^2}\left(1 + \frac{\sigma^2}{n\sigma_0^2}\right)\left[a - \frac{1}{1 + \frac{\sigma^2}{n\sigma_0^2}}\left(\frac{1}{n}\sum_{k=1}^{n} x_k + \frac{a_0\sigma^2}{n\sigma_0^2}\right)\right]^2\right\}, \tag{7.46}$$

donc, tout comme la densité de probabilité a priori, elle est normale, avec cette seule différence que les paramètres de la densité de probabilité a posteriori dépendent de l'échantillon d'apprentissage:

$$a_n = m_1\{a\} = \frac{1}{1 + \frac{\sigma^2}{n\sigma_0^2}}\left(\frac{1}{n}\sum_{k=1}^{n} x_k + \frac{a_0\sigma^2}{n\sigma_0^2}\right), \tag{7.47}$$

$$\sigma_n^2 = M_2\{a\} = \frac{\sigma^2}{n + \frac{\sigma^2}{\sigma_0^2}}. \tag{7.48}$$

Par conséquent, la fonction de vraisemblance $W_1(X \mid x_1, \ldots$ $\ldots, x_n;\ s_1)$ suit une loi normale de paramètres a_n et σ_n^2, de plus, la règle de classification de l'observation X en présence de l'échantillon d'apprentissage $x_1, \ldots, x_n$ s'obtient à partir de (7.45) et (7.45') en remplaçant a_0 par la valeur de la moyenne conditionnelle a_n donnée par (7.47) et σ_0^2 par la variance conditionnelle σ_n^2 donnée par (7.48). Par conséquent, la structure du dispositif optimal de détection reste inchangé, seuls changent lors de l'apprentissage les paramètres des blocs constitutifs.

Pour $n \to \infty$ on obtient à partir de (7.47) et (7.48)

$$a_n \sim \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \sigma_n^2 \to 0,$$

et, par conséquent, lorsque la taille de l'échantillon d'apprentissage augmente indéfiniment, le dispositif optimal de détection ressemble à un filtre linéaire adapté (le second terme du premier membre de (7.45) tendra vers zéro).

L'exemple étudié peut être généralisé au cas des valeurs corrélées du signal et du bruit également corrélé, en supposant toujours que le bruit est additif et indépendant du signal. Supposons que la moyenne vectorielle du bruit soit nulle et que sa matrice de corrélation soit égale à $\mathbf{M}$. La répartition multidimensionnelle a priori du signal est caractérisée par la moyenne vectorielle $\mathbf{a}_0$ et la matrice de corrélation $\mathbf{M}_0$. La répartition a priori du mélange additif du signal et du bruit qui sont indépendants est également normale de moyenne vectorielle $\mathbf{a}_0$ et de matrice de corrélation $\mathbf{M} + \mathbf{M}_0$. Soit $\mathbf{x}_1, \ldots, \mathbf{x}_n$ un échantillon vectoriel d'apprentissage appartenant au mélange du signal et du bruit. La répartition a posteriori du signal après apprentissage est également normale, mais caractérisée par une moyenne vectorielle et une matrice de corrélation (conditionnelles) un peu différentes. Elles peuvent être obtenues à partir de la relation de récurrence suivante, vraie pour la moyenne vectorielle $\mathbf{a}_n$ et la matrice de corrélation $\mathbf{M}_n$ conditionnelles de la loi normale [1]:

$$\mathbf{a}_n = \mathbf{M}(\mathbf{M}_{n-1} + \mathbf{M})^{-1}\mathbf{a}_{n-1} + \mathbf{M}_{n-1}(\mathbf{M}_{n-1} + \mathbf{M})^{-1}\mathbf{x}_n, \tag{7.49}$$

$$\mathbf{M}_n = \mathbf{M}(\mathbf{M}_{n-1} + \mathbf{M})^{-1}\mathbf{M}_{n-1}. \tag{7.50}$$

En vertu de (7.49) et (7.50) on a

$$\mathbf{a}_n = \frac{\mathbf{M}}{n}\left(\mathbf{M}_0 + \frac{\mathbf{M}}{n}\right)^{-1}\mathbf{a}_0 + \mathbf{M}_0\left(\mathbf{M}_0 + \frac{\mathbf{M}}{n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \tag{7.51}$$

$$\mathbf{M}_n = \frac{\mathbf{M}}{n}\left(\mathbf{M}_0 + \frac{\mathbf{M}}{n}\right)^{-1}\mathbf{M}_0. \tag{7.52}$$

Les formules (7.51) et (7.52) deviennent particulièrement simples lorsque $\mathbf{M} = \lambda\mathbf{M}_0$ où λ est un nombre positif. On a alors

$$\mathbf{a}_n = \frac{\lambda}{\lambda + n}\mathbf{a}_0 + \frac{n}{\lambda + n}\cdot\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \tag{7.53}$$

$$\mathbf{M}_n = \frac{1}{\lambda + n}\mathbf{M}. \tag{7.54}$$

Dans le cas unidimensionnel, (7.53) et (7.54) deviennent (7.47) et (7.48) respectivement avec $\lambda = \sigma^2/\sigma_0^2$.

La fonction de vraisemblance $W(X \mid X_{appr}; s_1)$ est dans le cas général une densité de probabilité multidimensionnelle normale de moyenne vectorielle $\mathbf{a}_n$ et de fonction de corrélation $\mathbf{M}_n$ données respectivement par les formules (7.51) et (7.52). Soulignons que seules les moyennes conditionnelles dépendent des échantillons d'apprentissage (plus exactement, de la moyenne échantillonnée), la matrice de corrélation conditionnelle ne dépendant que de la taille n de l'échantillon d'apprentissage.

7.3.3. Triage simple et méthode de découvertes. Nous allons maintenant passer à l'étude de la détection d'un signal inconnu noyé dans un bruit lors de l'apprentissage sans maître, lorsque l'échantillon d'apprentissage *n'est pas classé*. Ceci signifie que l'on ne sait pas d'avance si la valeur échantillonnée donnée appartient au mélange d'un signal et d'un bruit ou seulement à un bruit. Dans la construction de la règle de classification de l'observation X, d'après le critère du maximum de probabilité a posteriori on se heurte tout de suite à la difficulté de calculer la densité de probabilité a posteriori $W_1(a \mid X_{appr}; s_1)$ du paramètre pour un échantillon non classé $X_{appr} = (x_1, \ldots, x_n)$.

Il semble que dans ce cas le plus simple soit de composer toutes les suites possibles $x_1, \ldots, x_n$ dont chacun des termes x_i peut appartenir soit au mélange d'un signal et d'un bruit (état s_1) soit seulement au bruit (état s_0). Pour une suite quelconque $S_n^{(k)}$ d'états, lors de l'apprentissage on a

$$S_n^{(k)} = (s^{(1)}, \ldots, s^{(n)}),$$

où chacun des éléments $s^{(j)}$, $j = 1, \ldots, n$ peut être soit s_1, soit s_0, la densité de probabilité a posteriori $W(a \mid X_{appr}; S_n^{(k)})$ peut être calculée par la méthode du § 7.3.2, c'est-à-dire pour l'apprentissage avec maître. Le nombre des différentes suites pour l'échantillon d'apprentissage de dimension n est alors égal à 2^n. On a alors

$$W_1(a \mid X_{appr}) = \sum_{k=1}^{2^n} W_1(a \mid X_{appr}; S_n^{(k)}) P\{S_n^{(k)} \mid X_{appr}\}, \qquad (7.55)$$

où $P\{S_n^{(k)} \mid X_{appr}\}$ est la probabilité pour que dans l'échantillon d'apprentissage, la suite d'états $S_n^{(k)}$ puisse être réalisée. (On peut facilement calculer cette probabilité en supposant, par exemple, que l'apparition ou la non-apparition du signal lors de l'apprentissage soient indépendantes et que la probabilité a priori d'apparition du signal dans une observation quelconque soit connue.)

En vertu de (7.55) le dispositif optimal de détection lors de l'autoapprentissage sur un échantillon de taille n doit se composer de 2^n dispositifs, identiques à ceux qui ont été utilisés lors de l'apprentissage avec maître (filtre adapté *FA* et récepteur énergétique *RE*). Les signaux de sortie des dispositifs mentionnés sont multipliés par les coefficients de pondération $P\{S_n^{(k)} \mid X_{appr}\}$ puis les produits

sont sommés. Sur la figure 7.1 est représenté le schéma de principe d'un tel dispositif pour le cas où l'échantillon d'apprentissage est formé de deux éléments. Pour $n = 3$ le dispositif serait deux fois plus compliqué. Pour un échantillon de taille de l'ordre de quelques dizaines, la réalisation pratique d'un tel dispositif est plus que problématique. Il est évident que lorsque l'échantillon d'apprentissage est de taille n trop petite, les caractéristiques de classification sont mauvaises, et au fur et à mesure de l'augmentation de n elles tendent vers celles des algorithmes de statistique complètement connue des classes reconnaissables. Cependant, le dispositif optimal pour apprentissage sans maître devient de plus en plus compliqué, ceci à raison de 2^n.

Il a été montré dans [44] que pour trouver la densité de probabilité a posteriori du paramètre inconnu de la classe, on peut indiquer

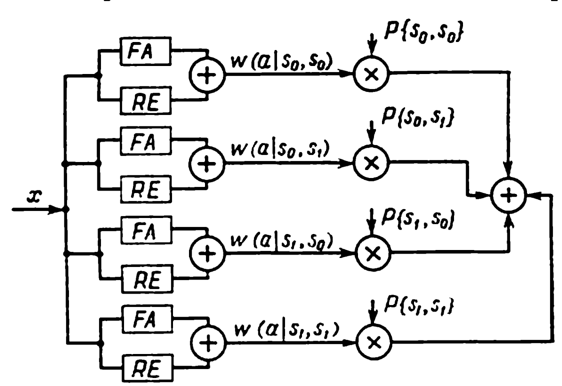


Fig. 7.1. Schéma d'un dispositif adaptatif de détection d'un signal inconnu travaillant en auto-apprentissage

l'algorithme de calcul fini de classification ne dépendant pas de la taille de l'échantillon d'apprentissage si et seulement si les observations d'apprentissage $x_1, \ldots, x_n$ sont caractérisées par une statistique suffisante de dimension finie. Pour le problème étudié ici de la détection d'un signal inconnu, la densité de probabilité d'un élément quelconque de l'échantillon d'apprentissage (sans maître) est [cf. (7.13)]:

$$w_1(x_i \mid a) = \frac{p}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-a)^2}{2\sigma^2}\right] + \frac{1-p}{\sigma\sqrt{2\pi}} \exp\left[-\frac{x_i^2}{2\sigma^2}\right], \quad (7.56)$$

où p est la probabilité de présence du signal. Dans [11] il est démontré que pour les densités de probabilité de la forme (7.56) il n'existe pas de statistique suffisante de dimension finie.

Ceci veut dire que, avec augmentation de la taille de l'échantillon d'apprentissage, le dispositif optimal de détection dans le cas d'auto-apprentissage doit devenir infiniment compliqué.

Nous allons étudier une possibilité de simplifier l'algorithme de classification dans le cas d'auto-apprentissage au prix d'une perte d'optimalité. Comme le dispositif optimal de détection d'un signal inconnu lors de l'apprentissage avec maître a une structure suffisamment simple, et que le dispositif analogue pour l'auto-apprentissage se complique beaucoup, dans ce dernier cas il semble naturel de deviner la suite des états (s_1 étant le mélange du signal et du bruit, s_0, le bruit pur) en passant d'un élément de l'échantillon d'apprentissage à l'autre, puis d'utiliser ce que l'on a deviné comme si c'était réellement exact (comme si c'était le « maître ») *). La structure du dispositif de détection devient alors identique à celle du dispositif optimal de détection avec maître, les résultats des découvertes lors de l'auto-apprentissage sont alors utilisés pour modifier les paramètres de ce dispositif conformément aux formules du § 7.3.2. Mais dans ce cas des erreurs sont inévitables et le dispositif de détection n'est plus optimal. Cependant, comme mentionné dans [44], pour qu'un tel dispositif de détection donne des résultats acceptables, il n'est pas indispensable que la suite devinée d'états soit exacte. Lorsque l'échantillon d'apprentissage est de taille n importante, il existe un sous-ensemble de l'ensemble de toutes les 2^n suites possibles d'états doué des deux propriétés suivantes: 1) la probabilité de présence de la suite réelle des états dans ce sous-ensemble est voisine de l'unité; 2) si la suite devinée est un élément de ce sous-ensemble, le dispositif de détection utilisant la méthode heuristique devient à la limite, pour $n \to \infty$, optimal. Malheureusement, on n'a pas jusqu'à présent découvert de méthodes efficaces de recherche des sous-ensembles doués des propriétés mentionnées.

Une variante de la méthode heuristique de découverte d'une suite d'états dans l'échantillon d'apprentissage est basée sur l'hypothèse selon laquelle la suite des *décisions* lors de l'auto-apprentissage est une suite de découvertes [41]. D'après cette méthode, dès le début, on attribue des valeurs aléatoires aux paramètres du dispositif de détection du signal inconnu. Chaque fois que le dispositif délivre une décision sur la présence d'un signal, les paramètres du dispositif changent conformément au signal reçu à ce même instant. Une méthode quelque peu modifiée a été utilisée dans [29].

7.3.4. Méthode adaptative de Bayes. L'application de la méthode générale de Bayes dont il était question dans les paragraphes précédents, exige la connaissance des répartitions a priori des paramètres inconnus. Au § 7.3.1 de ce chapitre nous avons étudié le cas où la forme de la répartition initiale a priori du paramètre de la classe changeait au cours de l'apprentissage avec maître. Est-il possible d'utiliser dans les problèmes de classification l'apprentissage avec

*) C'est pourquoi la méthode où il y a lieu de deviner est parfois appelée apprentissage avec maître réel, c'est-à-dire avec un maître pouvant se tromper.

maître afin d'être en mesure d'utiliser la méthode de Bayes sans répartition initiale a priori? La réponse affirmative à cette question est basée sur un théorème établi par S. Bernstein et R. Mises [14]. Donnons-en l'essentiel. Soit x_1 une valeur échantillonnée appartenant à une répartition de paramètre ϑ inconnu (aléatoire). La densité de probabilité a posteriori de ce paramètre est

$$W_1(\vartheta \mid x_1) = w_1(\vartheta) \frac{W_1(x_1 \mid \vartheta)}{\int\limits_{-\infty}^{\infty} w_1(\vartheta) W_1(x_1 \mid \vartheta)\, d\vartheta}, \tag{7.57}$$

où $w_1(\vartheta)$ est la densité de probabilité a priori du paramètre ϑ. Si l'on extrait la valeur échantillonnée suivante x_2, la fonction $W_1(\vartheta \mid x_1)$ peut être utilisée comme une nouvelle loi a priori pour le calcul de la densité de probabilité a posteriori, soit:

$$W_1(\vartheta \mid x_1, x_2) = W_1(\vartheta \mid x_1) \frac{W_1(x_2 \mid \vartheta)}{\int\limits_{-\infty}^{\infty} W_1(\vartheta \mid x_1) W_1(x_2 \mid \vartheta)\, d\vartheta} =$$

$$= w_1(\vartheta) \frac{W_1(x_1 \mid \vartheta) W_1(x_2 \mid \vartheta)}{\int\limits_{-\infty}^{\infty} w_1(\vartheta) W_1(x_1 \mid \vartheta) W_1(x_2 \mid \vartheta)\, d\vartheta}. \tag{7.58}$$

D'une manière analogue pour un échantillon de n éléments indépendants $x_1, \ldots, x_n$ on a

$$W_1(\vartheta \mid x_1, \ldots, x_n) = w_1(\vartheta) \frac{W_n(x_1, \ldots, x_n \mid \vartheta)}{\int\limits_{-\infty}^{\infty} w_1(\vartheta) W_n(x_1, \ldots, x_n \mid \vartheta)\, d\vartheta}, \tag{7.59}$$

où

$$W_n(x_1, \ldots, x_n \mid \vartheta) = \prod_{h=1}^{n} W_1(x_i \mid \vartheta).$$

Selon le théorème mentionné ci-dessus, si la densité de probabilité a priori $W_1(\vartheta)$ du paramètre ϑ est continue, au fur et à mesure de l'augmentation de la taille de l'échantillon la densité de probabilité a posteriori $W_1(\vartheta \mid x_1, \ldots, x_n)$ cesse de dépendre de la loi a priori. Ainsi, si seulement une densité de probabilité du paramètre a priori continue existe, pour un n suffisamment grand, il est plus ou moins indifférent quelle fonction $w_1(\vartheta)$ sera utilisée dans la formule (7.59). Ce théorème limite paraît avoir été à la base de la méthode adaptative de Bayes *) de classification proposée par Robbins [14], [16].

Soit $(x_1, \vartheta_1), \ldots, (x_n, \vartheta_n)$ une suite de couples de variables aléatoires indépendantes, tous les ϑ_i, $i = 1, \ldots, n$ étant soumis à la

*) Souvent la méthode de découverte ici est appelée méthode empirique de Bayes.

même loi a priori (inconnue) $w_1(\vartheta)$ et tous les x_i à la loi

$$W_1(x) = \int_{-\infty}^{\infty} W_1(x \mid \vartheta)\, w_1(\vartheta)\, d\vartheta.$$

Pour des x_i discrets on a

$$P\{x_i = x\} = \int_{-\infty}^{\infty} P\{x_i = x \mid \vartheta\}\, w_1(\vartheta)\, d\vartheta.$$

Si l'on adopte la décision γ selon laquelle l'observation X appartient à la loi $W_1(x \mid \vartheta)$, des pertes apparaissent, celles-ci étant données par une fonction non négative $\Pi(\gamma, \vartheta)$. Comme on ne connaît pas $w_1(\vartheta)$, l'algorithme du choix de la décision est établi d'après l'échantillon d'apprentissage $x_1, \ldots, x_n$ (les variables $\vartheta_1, \ldots, \vartheta_n$ restent toujours inconnues). Il est souhaitable qu'avec l'augmentation de la taille de l'échantillon d'apprentissage cet algorithme s'approche de celui de Bayes quand $w_1(\vartheta)$ est donné. L'algorithme satisfaisant à cette condition est appelé asymptotiquement optimal. Dans [17] on peut trouver une méthode de construction des algorithmes asymptotiquement optimaux. Nous allons nous limiter dans l'exposé de cette méthode au cas du problème à deux alternatives de la théorie des décisions statistiques avec loi de répartition a priori inconnue du paramètre ϑ, donnant la statistique des observations (sous sa forme générale, le problème est posé dans [17]).

Introduisons la notation

$$\Delta(x) = \int_{-\infty}^{\infty} [\Pi(\gamma_1, \vartheta) - \Pi(\gamma_0, \vartheta)]\, W_1(x \mid \vartheta)\, w_1(\vartheta)\, d\vartheta \quad (7.60)$$

et supposons que la fonction $\Delta_n(X)$ de l'observation X, dont la forme dépend de l'échantillon d'apprentissage $x_1, \ldots, x_n$, converge en probabilité vers $\Delta(X)$ pour $n \to \infty$. Dans ce cas on aura l'algorithme optimal asymptotique suivant: la décision γ_0 est adoptée si

$$\Delta_n(X) \geqslant 0, \quad (7.61)$$

et la décision γ_1 pour l'inégalité inverse de (7.61). (Dans le cas des répartitions discrètes il y a lieu de remplacer $W_1(x \mid \vartheta)$ par $P\{x_i = x \mid \vartheta\}$ dans (7.60).)

Considérons un exemple illustrant la possibilité de trouver une suite de fonctions $\Delta_n(x)$ convergeant en probabilité vers $\Delta(x)$. On vérifie l'hypothèse unilatérale H_0 selon laquelle le paramètre ϑ de la loi de Poisson d'une variable aléatoire n'est pas supérieur à la valeur donnée ϑ_0, c'est-à-dire $\vartheta \leqslant \vartheta_0$. Soit γ_0 la décision correspondant à l'acceptation de l'hypothèse H_0, et γ_1 la décision correspondant à son rejet. Introduisons la fonction de pertes $\Pi(\gamma_i, \vartheta)$ sous la

forme suivante:

$$\Pi(\gamma_0, \vartheta)=\begin{cases} 0, & \vartheta\leqslant\vartheta_0, \\ \vartheta-\vartheta_0, & \vartheta\geqslant\vartheta_0, \end{cases} \quad (7.62)$$

$$\Pi(\gamma_1, \vartheta)=\begin{cases} \vartheta_0-\vartheta, & \vartheta\leqslant\vartheta_0, \\ 0, & \vartheta\geqslant\vartheta_0. \end{cases} \quad (7.62')$$

Pour la loi de Poisson d'une variable aléatoire discrète on a

$$P\{x=x_i \mid \vartheta\}=\frac{\vartheta^x}{x!}e^{-\vartheta}, \quad \vartheta>0, \; x=0, 1, 2, \ldots \quad (7.63)$$

Portant (7.63) dans (7.60) au lieu de $W_1(x \mid \vartheta)$ on obtient compte tenu de (7.62) et (7.62′)

$$\Delta(x)=\int_0^\infty (\vartheta_0-\vartheta)\frac{\vartheta^x}{x!}e^{-\vartheta}w_1(\vartheta)\,d\vartheta=\vartheta_0F_1(x)-(x+1)F_1(x+1), \quad (7.64)$$

où

$$F_1(x)=\int_0^\infty \frac{\vartheta^x}{x!}e^{-\vartheta}w_1(\vartheta)\,d\vartheta=P\{x_i=x\}. \quad (7.65)$$

Introduisons le compteur de coïncidences

$$\nu(x, y)=\begin{cases} 1, & x=y, \\ 0, & x\neq y \end{cases} \quad (7.66)$$

et considérons la fonction suivante de x dépendant de l'échantillon d'apprentissage $x_1, \ldots, x_n$:

$$u_n(x)=\frac{1}{n}\sum_{i=1}^n \nu(x, x_i). \quad (7.67)$$

La somme dans le second membre de (7.67) est égale au nombre des valeurs échantillonnées d'apprentissage qui sont exactement égales à x, par conséquent, $u_n(x)$ est la répartition empirique de x convergeant en probabilité pour $n\to\infty$ vers $F_1(x)$ [cf. (7.65)]. La fonction

$$\Delta_n(x)=\vartheta_0u_n(x)-(x+1)\,u_n(x+1) \quad (7.68)$$

tend en probabilité vers $\Delta(x)$ définie conformément à (7.64). On en déduit l'algorithme optimal asymptotique suivant: à la base de l'échantillon d'apprentissage $x_1, \ldots, x_n$ on adopte la décision γ_0 (acceptation de l'hypothèse H_0) si

$$\frac{\vartheta_0}{n}\sum_{i=1}^n \nu(X, x_i)-\frac{X+1}{n}\sum_{i=1}^n \nu(X+1, x_i)\geqslant 0, \quad (7.69)$$

et la décision γ_1 (rejet de l'hypothèse H_0) pour l'inégalité inverse de (7.69).

7.4. APPROXIMATION STOCHASTIQUE

7.4.1. Les zéros et les extréma de la fonction de régression. Supposons qu'à chaque valeur du paramètre ϑ corresponde une variable aléatoire x dont la fonction de répartition est $F_1(x \mid \vartheta)$ avec $m_1\{x \mid \vartheta\} = m(\vartheta)$. Proposons-nous de trouver la solution de l'équation de régression

$$m(\vartheta) = 0 \tag{7.70}$$

en supposant que celle-ci ait une seule racine et que les fonctions $F_1(x \mid \vartheta)$ et $m(\vartheta)$ soient données. Robbins et Monroe [13] ont proposé une méthode itérative appelée *approximation stochastique* permettant de construire les estimations de la racine cherchée à l'aide de l'échantillon d'apprentissage $x_1, \ldots, x_n$, pour chaque élément x_k duquel on a

$$m_1\{x_k \mid \hat{\vartheta}_k\} = m(\hat{\vartheta}_k).$$

L'estimation $\hat{\vartheta}_{n+1}$ de la racine cherchée peut être trouvée à partir de l'estimation $\hat{\vartheta}_n$ d'après la valeur échantillonnée d'apprentissage x_n, ceci à l'aide de la relation

$$\hat{\vartheta}_{n+1} = \hat{\vartheta}_n + a_n x_n, \quad n \geqslant 1, \tag{7.71}$$

$\hat{\vartheta}_1 = \vartheta_1$ étant une constante quelconque. Si les coefficients a_n satisfont aux conditions

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty, \tag{7.72}$$

l'estimation $\hat{\vartheta}_{n+1}$ tend en probabilité, avec augmentation infinie de la taille n de l'échantillon, vers la racine de l'équation (7.70). Notons, par exemple, que la série harmonique $\left(a_n = \frac{1}{n}\right)$ satisfait aux conditions (7.72).

On peut utiliser la même méthode itérative pour trouver l'extrémum de la fonction unimodale de régression $m(\vartheta)$. Cette méthode, proposée par Kiffer et Volfovitz [13], permet de trouver l'estimation $\hat{\vartheta}_{n+1}$ de la valeur extrémale de la fonction de régression d'après l'estimation précédente $\hat{\vartheta}_n$ ainsi que les valeurs échantillonnées x_{2n} et x_{2n-1} :

$$\hat{\vartheta}_{n+1} = \hat{\vartheta}_n + a_n \frac{x_{2n} - x_{2n-1}}{c_n}, \quad n \geqslant 1, \tag{7.73}$$

les échantillons x_{2n} et x_{2n-1} étant indépendants et correspondant aux valeurs $\hat{\vartheta}_n + c_n$ et $\hat{\vartheta}_n - c_n$ du paramètre. Comme précédemment, la valeur initiale ϑ_1 est arbitraire. Si les coefficients a_n et c_n satisfont aux conditions

$$a_n > 0, \; c_n > 0, \; c_n \to 0 \text{ pour } n \to \infty, \tag{7.74}$$

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n c_n < \infty, \quad \sum_{n=1}^{\infty} \left(\frac{a_n}{c_n}\right)^2 < \infty, \tag{7.74'}$$

avec augmentation illimitée de la taille n de l'échantillon d'apprentissage l'estimation $\hat{\vartheta}_{n+1}$ calculée conformément à (7.73) tend en probabilité vers la valeur extrémale de la fonction de régression.

Si l'on impose des conditions supplémentaires aux fonctions $m(\vartheta)$, les estimations $\hat{\vartheta}_{i+1}$ données par (7.71) et (7.73) peuvent converger en moyenne quadratique respectivement vers zéro et vers l'extrémum de la fonction de régression. Les méthodes examinées peuvent être généralisées au cas multidimensionnel (voir [13]).

Tout ce qui a été dit se rapporte également à l'estimation du zéro ou de l'extrémum de la fonction $m_1 \{f(x) \mid \vartheta\}$ où $f(x)$ est une fonction donnée de la variable aléatoire x dont la loi de répartition est donnée par $F_1(x \mid \vartheta)$. Dans ce cas dans les formules (7.71) et (7.73) il y a lieu de remplacer les valeurs échantillonnées d'apprentissage x_k par $f(x_k)$.

L'inconvénient des méthodes stochastiques d'approximation est qu'à l'heure actuelle il n'existe pas de considérations générales, plus ou moins fondées, permettant de choisir les coefficients a_n et c_n dans (7.71) et (7.73), pour assurer une convergence rapide des itérations, ni de moyens permettant d'apprécier la précision des estimations et d'arrêter l'itération à un certain moment. En revanche, les itérations dans les méthodes d'approximation stochastique sont relativement simples et faciles à réaliser.

7.4.2. Estimation des paramètres et des fonctions de répartition. L'approximation stochastique peut être utilisée pour estimer les paramètres et les fonctions de répartition inconnus des classes [20]. A titre d'exemple particulièrement simple examinons l'estimation d'une moyenne inconnue. Supposons que le paramètre dans (7.70) soit une moyenne inconnue. En posant alors $f(x) = x - \vartheta_0$ on obtient l'équation de régression

$$m_1 \{x - \vartheta_0 \mid \vartheta\} = 0,$$

et à partir de (7.71) on trouve la relation de récurrence pour l'estimation de la moyenne inconnue (racine de l'équation de régression)

$$\hat{\vartheta}_{n+1} = \hat{\vartheta}_n + a_n (x_n - \hat{\vartheta}_n) = \hat{\vartheta}_n (1 - a_n) + a_n x_n. \qquad (7.75)$$

Pour $a_n = \frac{1}{n}$, on a à partir de (7.75)

$$\hat{\vartheta}_{n+1} = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad (7.75')$$

c'est-à-dire que l'estimation de la moyenne est égale à la moyenne arithmétique des valeurs échantillonnées d'apprentissage. Soit σ^2 la variance inconnue de la variable aléatoire x de moyenne nulle, en posant $f(x) = x^2 - \sigma_0^2$ on obtient alors l'équation de régression:

$$m_1 \{x^2 - \sigma_0^2 \mid \sigma^2\} = 0$$

et l'algorithme de l'estimation de la variance correspondant s'écrit

comme suit:

$$\hat{\sigma}^2_{n+1} = \hat{\sigma}^2_n + a_n (x^2_n - \hat{\sigma}^2_n) = \hat{\sigma}^2_n (1 - a_n) + a_n x^2_n. \qquad (7.76)$$

Pour $a_n = \frac{1}{n}$ on a à partir de (7.76)

$$\hat{\sigma}^2_{n+1} = \frac{1}{n} \sum_{i=1}^{n} x_i^2. \qquad (7.76')$$

On peut également utiliser l'approximation stochastique pour estimer la densité de probabilité inconnue $w_1(x)$ d'après les échantillons d'apprentissage caractérisés par cette densité, en la développant en série suivant les fonctions orthogonales $Q_k(x)$ [cf. (2.120), tome I]. L'estimation de la densité de probabilité inconnue se réduit alors à l'estimation des coefficients du développement

$$c_n^{(k)} = m_1 \{Q_k(x)\} \qquad (7.77)$$

ce qu'on obtient à partir de la relation

$$\hat{c}^{(k)}_{n+1} = \hat{c}^{(k)}_n + a_n^{(k)} [Q_k(x_n) - \hat{c}^{(k)}_n], \; k = 1, 2, \ldots \qquad (7.78)$$

Pour $a^{(k)}_n = \frac{1}{n}$ on a

$$\hat{c}^{(k)}_{n+1} = \frac{1}{n} \sum_{i=1}^{n} Q_k(x_i). \qquad (7.78')$$

7.4.3. Discrimination de deux signaux inconnus. Pour illustrer la procédure itérative (7.73) nous allons étudier le problème de la discrimination de deux signaux s_0 et s_1 dont les densités de probabilité $w_1(x \mid s_0)$ et $w_1(x \mid s_1)$ sont inconnues. L'algorithme du choix d'une décision consiste à comparer l'observation X avec un seuil ϑ. Si $X < \vartheta$ l'observation est rapportée au signal s_0, et si $X \geqslant \vartheta$, au signal s_1. En utilisant les échantillons d'apprentissage on choisit le seuil ϑ de façon à minimiser le risque moyen.

En vertu de (1.18) le risque moyen est égal dans ce cas à

$$R(\vartheta) = \int_{\vartheta}^{\infty} [q\Pi_{01} w_1(x \mid s_0) + p\Pi_{11} w_1(x \mid s_1)]\, dx + \int_{-\infty}^{\vartheta} [q\Pi_{00} w_1(x \mid s_0) + p\Pi_{10} w_1(x \mid s_1)]\, dx, \qquad (7.79)$$

où p et q sont les probabilités a priori d'apparition des signaux s_1 et s_0, et Π_{ij} $(i, j = 0, 1)$ les éléments de la matrice de coût.

Introduisons la fonction de décision (cf. § 1.2.1)

$$\Phi(x, \vartheta) = \begin{cases} 0, & x < \vartheta, \\ 1, & x \geqslant \vartheta, \end{cases} \qquad (7.80)$$

l'expression du risque moyen s'écrit alors

$$R(\vartheta) = \int_{-\infty}^{\infty} \{\Phi(x, \vartheta)[q\Pi_{01}w_1(x \mid s_0) + p\Pi_{11}w_1(x \mid s_1)] +$$
$$+ [1 - \Phi(x, \vartheta)][q\Pi_{00}w_1(x \mid s_0) + p\Pi_{10}w_1(x \mid s_1)]\} dx. \quad (7.81)$$

En vertu de (7.81), le risque moyen $R(\vartheta)$ est une fonction de régression pour une variable aléatoire discrète pouvant prendre quatre valeurs

$$y = f(x) = \Pi_{ij}, \ i, j = 0, 1, \quad (7.82)$$

ici x est une valeur échantillonnée extraite de la répartition du signal s_i et

$$\Phi(x, \vartheta) = j. \quad (7.82')$$

Ainsi,

$$R(\vartheta) = m_1 \{f(x) \mid \vartheta\}, \quad (7.83)$$

et le problème de la détermination du seuil optimal ϑ^* pour la classification de l'observation revient au calcul du minimum de la fonction de régression (7.83), les lois de répartition des signaux étant inconnues.

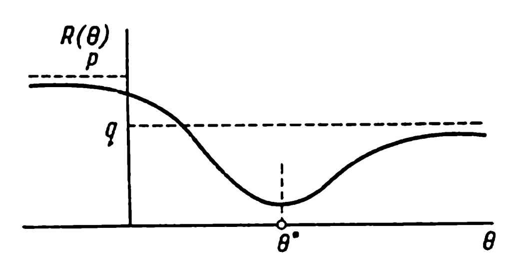


Fig. 7.2. Fonction du risque moyen

Laissant de côté l'étude de la fonction de régression admettant plusieurs extréma, limitons-nous au cas d'un minimum unique, comme on peut le voir sur la figure 7.2 (pour $\Pi_{10} = \Pi_{01} = 1$ et $\Pi_{00} = \Pi_{11} = 0$).

Pour estimer le seuil optimal ϑ^* on peut utiliser l'algorithme (7.73)

$$\hat{\vartheta}_{n+1} = \hat{\vartheta}_n + a_n \frac{y_{2n} - y_{2n-1}}{c_n}, \quad (7.84)$$

où $y_{2n} = \Pi_{ij}$ lorsque x_{2n} est la valeur échantillonnée extraite de la répartition du signal s_i et $\Phi(x_{2n}, \hat{\vartheta}_n + c_n) = j$ et $y_{2n-1} = \Pi_{ij}$ lorsque c'est x_{2n-1} qui appartient à la répartition du signal s_i et $\Phi(x_{2n-1}, \hat{\vartheta}_n - c_n) = j$.

Lors du premier pas l'estimation $\hat{\vartheta}_1$ est une constante arbitraire et les coefficients a_n et c_n satisfont aux conditions (7.74), (7.74').

Lorsque la taille n des échantillons d'apprentissage augmente indéfiniment, l'estimation $\hat{\vartheta}_{n+1}$ tend en probabilité vers le seuil optimal ϑ^*, pour lequel le risque moyen $R(\vartheta^*)$ est minimal. En adoptant des hypothèses supplémentaires sur la fonction $R(\vartheta)$ on peut également assurer la convergence en moyenne quadratique (voir [26]).

ANNEXES

ANNEXE I

Loi normale

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, dt.$$

$$\varphi(x) = F'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	$F(x)$	$\varphi(x)$	x	$F(x)$	$\varphi(x)$
0,0	0,50000	0,39894	2,0	0,97725	0,05399
0,1	0,53983	0,39695	2,1	0,98214	0,04398
0,2	0,57926	0,39104	2,2	0,98610	0,03547
0,3	0,61791	0,39139	2,3	0,98928	0,02833
0,4	0,65542	0,36827	2,4	0,99180	0,02239
0,5	0,69146	0,35207	2,5	0,99379	0,01753
0,6	0,72575	0,33322	2,6	0,99534	0,01358
0,7	0,75804	0,31225	2,7	0,99653	0,01042
0,8	0,78814	0,28969	2,8	0,99744	0,00792
0,9	0,81594	0,26609	2,9	0,99813	0,00595
1,0	0,84134	0,24197	3,0	0,99865	0,00443
1,1	0,86433	0,21785	3,1	0,99903	0,00327
1,2	0,88493	0,19419	3,2	0,99931	0,00238
1,3	0,90320	0,17137	3,3	0,99952	0,00172
1,4	0,91924	0,14973	3,4	0,99966	0,00123
1,5	0,93319	0,12952	3,5	0,99977	0,00087
1,6	0,94520	0,11092	3,6	0,99984	0,00061
1,7	0,95543	0,09405	3,7	0,99989	0,00042
1,8	0,96407	0,07895	3,8	0,99993	0,00029
1,9	0,97128	0,06562	3,9	0,99995	0,00020

$$K(r, h) = \frac{1}{2\pi\sqrt{1-r^2}} \int_h^\infty \int_h^\infty e^{-\frac{u^2+v^2-2ruv}{2(1-r^2)}} du\, dv$$

r \ h	0	0.5	1.0	1.5	2.0
0	0,250000	0,095195	0,025171	0,004463	0,000518
0,1	0,265942	0,107758	0,031320	0,006334	0,000872
0,2	0,282047	0,120715	0,038069	0,008611	0,001370
0,3	0,298493	0,134179	0,045485	0,011330	0,002047
0,4	0,315495	0,148306	0,053563	0,014542	0,002921
0,5	0,333333	0,163320	0,062514	0,018323	0,004053
0,6	0,352416	0,179560	0,075526	0,022794	0,005500
0,7	0,373407	0,197602	0,083979	0,028166	0,007362
0,8	0,397583	0,218566	0,097637	0,034856	0,009825
0,9	0,428214	0,245325	0,115490	0,043948	0,013361
1,0	0,500000	0,308538	0,158655	0,066807	0,022750

ANNEXE II

Calcul des intégrales

Soit une intégrale du type

$$K = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\varphi(x,\, y)}\, dx\, dy, \tag{1}$$

où

$$\varphi\,(x,\, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} \tag{2}$$

est une forme quadratique définie positive.

L'expression (2) est l'équation d'une surface du second degré, d'un ellipsoïde. Par rotation des axes et translation de l'origine des coordonnées on peut mettre cette équation sous la forme canonique

$$\Phi\,(u,\, v) = \lambda_1 u^2 + \lambda_2 v^2 + c, \tag{3}$$

qui en plus du terme constant ne contient que les carrés des variables u et v.

Avec ce changement de variables l'intégrale (1) s'écrit

$$K = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\Phi(u,\, v)}\, D\, du\, dv,$$

où D est le jacobien de la transformation.

Comme cette transformation des coordonnées se réduit à une translation de l'origine et à une rotation des axes, on a $D = 1$.

Ainsi

$$K=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2}(\lambda_1u^2+\lambda_2v^2+c)}\,du\,dv=e^{-\frac{c}{2}}\int_{-\infty}^{\infty}e^{-\frac{\lambda_1u^2}{2}}\,du\int_{-\infty}^{\infty}e^{-\frac{\lambda_2v^2}{2}}\,dv.$$

Mais

$$\int_{-\infty}^{\infty}e^{-\frac{\lambda z^2}{2}}\,dz=\sqrt{\frac{2\pi}{\lambda}},$$

on a

$$K=\frac{2\pi}{\sqrt{\lambda_1\lambda_2}}e^{-\frac{c}{2}}. \tag{4}$$

Ainsi le calcul de l'intégrale (1) revient à trouver les grandeurs λ_1, λ_2 et c. Dans la théorie des formes quadratiques on montre que λ_1 et λ_2 sont les racines de l'équation caractéristique

$$\lambda^2-I_1\lambda+I_2=0, \tag{5}$$

dont les coefficients s'expriment en fonction des coefficients de la forme quadratique (2)

$$I_1=a_{11}+a_{22},\quad I_2=a_{11}a_{22}-a_{12}^2, \tag{6}$$

et la constante c est donnée par la formule

$$c=\frac{I_3}{I_2}, \tag{7}$$

où

$$I_3=\begin{vmatrix}a_{11} & a_{12} & a_{13}\\ a_{12} & a_{22} & a_{23}\\ a_{13} & a_{23} & a_{33}\end{vmatrix}. \tag{8}$$

La formule (4) ne contient que le produit $\lambda_1\lambda_2$ des racines de l'équation (5) égal au terme constant I_2. L'intégrale cherchée est donc

$$K=\frac{2\pi}{\sqrt{I_2}}e^{-\frac{I_3}{2I_2}}. \tag{9}$$

Nous allons utiliser la formule (9) pour le calcul de la fonction caractéristique bidimensionnelle de la loi normale. En faisant préalablement le changement de variables d'intégration $x=\frac{x_1-a_1}{\sigma_1}$ et $y=\frac{x_2-a_2}{\sigma_2}$, on peut alors écrire $\Theta_2(v_1, v_2)$ comme suit

$$\Theta_2(v_1, v_2)=\frac{1}{2\pi\sqrt{1-r^2}}e^{i(a_1v_1+a_2v_2)}\times$$
$$\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{x^2-2rxy+y^2}{2(1-r^2)}+iv_1\sigma_1x+iv_2\sigma_2y\right]dx\,dy. \tag{10}$$

L'intégrale dans (10) est un cas particulier de (1) pour

$$a_{11}=a_{22}=\frac{1}{1-r^2}, \quad a_{12}=-\frac{r}{1-r^2}, \quad a_{13}=-i\sigma_1 v_1, \quad a_{23}=-i\sigma_2 v_2, \quad a_{33}=0.$$

On obtient à partir de (6) et (8)

$$I_2=\frac{1}{(1-r^2)^2}-\frac{r^2}{(1-r^2)^2}=\frac{1}{1-r^2}, \qquad I_3=\frac{2r\sigma_1\sigma_2 v_1 v_2+\sigma_1^2 v_1^2+\sigma_2^2 v_2^2}{1-r^2}.$$

On a alors

$$\Theta_2(v_1, v_2)=\frac{1}{2\pi\sqrt{1-r^2}}\, e^{i(a_1 v_1+a_2 v_2)}\times$$
$$\times\, 2\pi\sqrt{1-r^2}\exp\left[-\frac{1}{2}(\sigma_1^2 v_1^2+2r\sigma_1\sigma_2 v_1 v_2+\sigma_2^2 v_2^2)\right],$$

ce qui en fait ne diffère en rien de (3.123) pour $n = 2$.

Soit maintenant une intégrale du type

$$K_{mn}(\alpha)=\int_0^\infty\int_0^\infty u_1^m u_2^n \exp\left[-(u_1^2+u_2^2+2u_1u_2\cos\alpha)\right] du_1\, du_2. \quad (11)$$

Pour $n \geqslant m$ l'intégrale (11) peut être obtenue par dérivation successive par rapport au paramètre de l'intégrale

$$K_{0r}(\alpha)=\int_0^\infty\int_0^\infty u_2^r \exp\left[-(u_1^2+u_2^2+2u_1u_2\cos\alpha)\right] du_1\, du_2. \quad (12)$$

Pour le calcul de l'intégrale (12) il faut écrire l'exposant de l'exponentielle sous la forme d'un carré et faire le changement de variables d'intégration

$$v_1=u_1+u_2\cos\alpha, \quad v_2=u_2\sin\alpha.$$

On obtient alors

$$K_{0r}(\alpha)=\frac{1}{\sin^{r+1}\alpha}\int_0^\infty\int_{v_2\,\mathrm{ctg}\,\alpha}^\infty v_2^r e^{-(v_1^2+v_2^2)}\, dv_1\, dv_2$$

et, en passant aux coordonnées polaires, on trouve

$$K_{0r}(\alpha)=\frac{1}{\sin^{r+1}\alpha}\int_0^\alpha\int_0^\infty(\rho\sin\vartheta)^r e^{-\rho^2}\rho\, d\rho\, d\vartheta.$$

Les intégrales sur ρ et ϑ se séparent. Compte tenu de

$$\int_0^\infty \rho^{r+1}e^{-\rho^2}\, d\rho=\frac{1}{2}\Gamma\left(\frac{r}{2}+1\right),$$

on obtient

$$K_{0r}(\alpha)=\frac{\Gamma\left(\frac{r}{2}+1\right)}{2\sin^{r+1}\alpha}\int_0^\alpha \sin^r\vartheta\, d\vartheta. \quad (13)$$

En vertu de (13) pour $r = 0$ on a

$$K_{00}(\alpha) = \frac{\alpha}{2\sin\alpha}, \tag{14}$$

$$K_{11}(\alpha) = \frac{1}{2\sin\alpha}\frac{dK_{00}(\alpha)}{d\alpha} = \frac{1-\alpha\operatorname{ctg}\alpha}{4\sin^2\alpha}. \tag{15}$$

Pour $h = 0$ l'intégrale (10.86) est

$$q_1(\tau, 0) = \frac{D_{22}\sqrt{1-r^2}}{2\pi\omega_1(1-R^2)^{3/2}}\int_0^\infty\int_0^\infty v_1v_2 e^{-(v_1^2+v_2^2-2rv_1v_2)}\,dv_1\,dv_2 =$$

$$= \frac{4D_{22}(1-r^2)^{3/2}}{2\pi\omega_1(1-R^2)^{3/2}}K_{11}(\alpha) = \frac{D_{22}}{2\pi\omega_1}\left(\frac{1-r^2}{1-R^2}\right)^{3/2}\frac{1-\alpha\operatorname{ctg}\alpha}{\sin^2\alpha}, \tag{16}$$

avec

$$\cos\alpha = -r. \tag{17}$$

La substitution de (17) dans (16) conduit à (10.89).

On peut écrire (8.69) sous une forme analogue après les substitutions

$$r_1 = u_1\sigma\sqrt{2(1-R_0^2)}, \quad r_2 = u_2\sigma\sqrt{2(1-R_0^2)},$$

$$W_2(\vartheta_1, \vartheta_2, r) = \frac{1-R_0^2}{\pi^2}\int_0^\infty\int_0^\infty u_1u_2 e^{-(u_1^2+u_2^2-2u_1u_2y)}\,du_1\,du_2 =$$

$$= \frac{1-R_0^2}{\pi^2}K_{11}(\alpha) = \frac{1-R_0^2}{4\pi^2}\frac{1-\alpha\operatorname{ctg}\alpha}{\sin^2\alpha}, \tag{18}$$

avec de plus

$$-\cos\alpha = y = R_c\cos(\vartheta_2-\vartheta_1) + R_s\cos(\vartheta_2-\vartheta_1). \tag{19}$$

En portant (19) dans (18) on aboutit à (8.70).

ANNEXE III

Fonction delta

Par définition, la fonction delta $\delta(t-t_0)$ de tout paramètre réel t_0 est nulle pour $t \neq t_0$ et illimitée pour $t = t_0$

$$\delta(t-t_0) = \begin{cases} 0, & t \neq t_0, \\ \infty, & t = t_0. \end{cases} \tag{1}$$

L'intégrale de cette fonction est égale à

$$\int_a^b \delta(t-t_0)\,dt = \begin{cases} 1, & a < t_0 < b, \\ \frac{1}{2}, & t_0 = a \text{ ou } t_0 - b, \\ 0, & t_0 > a \quad t_0 > b. \end{cases} \tag{2}$$

En toute rigueur la fonction delta peut être obtenue comme la fonction limite d'une famille de fonctions continues uniparamétriques *). On peut donner de nombreux exemples de telles familles. Citons à titre d'exemple une famille de densités de probabilité de la loi normale de moyenne a constante et de moyenne quadratique σ variable.

Soit une autre famille

$$\varphi(t, \lambda) = \frac{\lambda}{\pi(\lambda^2 t^2 + 1)},$$

qui pour $\lambda \to \infty$ donne la fonction delta. Notre troisième exemple sera un ensemble $s(t, \tau)$ d'impulsions rectangulaires d'aire unité, de durée τ et de hauteur $\frac{1}{\tau}$

$$s(t, \tau) = \begin{cases} \frac{1}{\tau}, & t_0 < t < t_0 + \tau, \\ 0, & t < t_0, \ t > t_0 + \tau. \end{cases} \tag{3}$$

En faisant tendre vers zéro la durée de l'impulsion, on obtient à la limite une fonction delta, soit

$$\delta(t - t_0) = \lim_{\tau \to 0} s(t, \tau). \tag{4}$$

La convolution de la fonction delta avec une fonction quelconque $f(t)$, limitée et continue au point t_0, est douée de la propriété remarquable suivante:

$$\int_a^b f(t)\,\delta(t - t_0)\,dt = \begin{cases} f(t_0), & a < t_0 < b, \\ \frac{1}{2} f(t_0), & t_0 = a \text{ ou } t_0 = b, \\ 0, & t_0 < a, \ t_0 > b. \end{cases} \tag{5}$$

Si au point $t = t_0$ la fonction $f(t)$ a une discontinuité (de première espèce), on a

$$\int_a^b f(t)\,\delta(t - t_0)\,dt = \frac{1}{2}\left[f(t_0 +) - f(t_0 -)\right], \qquad a < t_0 < b, \tag{5'}$$

*) La fonction delta est un exemple des fonctions *généralisées*, définies comme les limites des suites de fonctions continues. On peut trouver la théorie de ces fonctions dans l'ouvrage de I. Gelfand et G. Schilov. *Les fonctions généralisées et leurs propriétés*. Physmathguiz, Moscou, 1958.

où $f(t_0+)$ et $f(t_0-)$ sont les valeurs de $f(t)$ à droite et à gauche du point de coupure.

La propriété exprimée par la formule (5) peut être appelée propriété de *filtrage* de la fonction delta. En effet, la fonction delta agit comme un filtre; en multipliant une fonction arbitraire $f(t)$ par $\delta(t-t_0)$ et en intégrant sur t on choisit une des valeurs de cette fonction, soit $f(t_0)$, c'est-à-dire la valeur qui correspond au zéro de l'argument de la fonction delta ($t-t_0=0$). Pour démontrer la formule (5) il suffit de substituer sous le signe de l'intégrale à $\delta(t-t_0)$ une fonction quelconque qui en est l'approximation et de passer ensuite à la limite.

Notons que la fonction delta $\delta(x-x_0)$ a la dimension de la grandeur $\frac{1}{x}$.

Cherchons maintenant le spectre (transformée de Fourier) de la fonction delta. En utilisant la propriété de filtrage on a

$$\int_{-\infty}^{\infty} \delta(t-t_0)\, e^{-i\omega t}\, dt = e^{-i\omega t_0}. \tag{6}$$

Si $t_0=0$, en vertu de (6) le spectre de $\delta(t)$ est *uniforme pour toutes les fréquences* d'intensité unité. Le spectre de la demi-somme de deux fonctions delta $\frac{1}{2}[\delta(t+t_0)+\delta(t-t_0)]$ en vertu de (6) est égal à $\frac{1}{2}(e^{i\omega t_0}+e^{-i\omega t_0})=\cos\omega t_0$.

Par transformation de Fourier inverse on obtient

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\, d\omega = \frac{1}{\pi}\int_{0}^{\infty} \cos\omega t\, d\omega = \delta(t), \tag{7}$$

$$\frac{1}{2}\int_{-\infty}^{\infty} e^{i\omega t}\cos\omega t_0\, d\omega = \frac{1}{\pi}\int_{0}^{\infty} \cos\omega t\cos\omega t_0\, d\omega = \frac{1}{2}[\delta(t+t_0)+\delta(t-t_0)]. \tag{7'}$$

Par suite de la symétrie de l'intégrale de Fourier les variables t et ω dans les formules (6) et (7) peuvent être interchangées.

Les dérivées des fonctions delta se définissent comme les limites des dérivées correspondantes des fonctions d'approximation. Ainsi, par exemple, en tant que fonction d'approximation on utilise des densités de probabilité de la loi normale pour $\sigma\to 0$, la n-ième dérivée de la fonction delta se définit comme suit

$$\delta^{(n)}(t) = \lim_{\sigma\to 0}\left\{\frac{1}{\sigma\sqrt{2\pi}}\frac{d}{dt^n}e^{-\frac{t^2}{2\sigma^2}}\right\}. \tag{8}$$

Tout comme la fonction delta, ses dérivées sont égales à zéro pour $t\neq 0$. L'allure des dérivées pour $t=0$ est compliquée. Ainsi,

par exemple, la dérivée première de la fonction delta

$$\delta'(t) = -\frac{t}{\sqrt{2\pi}} \lim_{\sigma\to 0} \frac{1}{\sigma^3} e^{-\frac{t^2}{2\sigma^2}}$$

est égale à $+\infty$ lorsqu'on s'approche de l'origine des coordonnées à gauche ($t = 0$ —) et à $-\infty$ lorsqu'on s'approche à droite ($t = 0 +$). Au voisinage de $t = 0$, $\delta'(t)$ se conduit à peu près comme t^{-1}.

Les dérivées de la fonction delta possèdent également les propriétés de filtrage. La convolution d'une dérivée d'ordre n de la fonction delta avec une fonction quelconque dont la dérivée d'ordre n est continue au point t_0 est égale à

$$\int_{-\infty}^{\infty} f(t)\, \delta^{(n)}(t - t_0)\, dt = (-1)^{(n)} f^{(n)}(t_0). \tag{9}$$

Si la dérivée $f^{(n)}(t)$ a une discontinuité (de première espèce) au point t_0, on a

$$\int_{-\infty}^{\infty} f(t)\, \delta^{(n)}(t-t_0)\, dt = \frac{(-1)^n}{2} [f^{(n)}(t_0+) + f(t_0-)]. \tag{9'}$$

Cherchons maintenant le spectre (transformée de Fourier) de la dérivée d'une fonction delta. En utilisant (9) on obtient

$$\int_{-\infty}^{\infty} \delta^{(n)}(t-t_0)\, e^{-i\omega t}\, dt = \left(\frac{d^n e^{-i\omega t}}{dt^n}\right)_{t=t_0} = (-i\omega)^n e^{-i\omega t_0}. \tag{10}$$

Si $t_0 = 0$, en vertu de (10) le spectre de $\delta^{(n)}(t)$ est égal à $(-i\omega)^n$.

ANNEXE IV

Systèmes orthogonaux de fonctions

Deux fonctions $f(x)$ et $g(x)$ définies sur l'intervalle (a, b) sont dites *orthogonales* par rapport à la fonction de pondération $\varphi(x)$ si l'on a

$$\int_a^b \varphi(x) f(x) g(x)\, dx = 0. \tag{1}$$

Le système de fonctions $f_1(x), f_2(x), \ldots, f_n(x)$ est dit *orthogonal* par rapport au poids $\varphi(x)$ si deux fonctions quelconques du système sont orthogonales, c'est-à-dire si

$$\int_a^b \varphi(x) f_i(x) f_j(x)\, dx = 0,\ i \neq j. \tag{2}$$

La quantité

$$\int_a^b \varphi(x)\, f_i^2(x) = a_i \tag{3}$$

est appelée *norme* de la fonction $f_i(x)$ par rapport au poids $\varphi(x)$. Si pour toute fonction satisfaisant à la condition (2) on a $a_i = 1$, le système de fonctions est dit *orthonormé*.

A titre d'exemple particulièrement simple de fonctions orthogonales on peut citer les fonctions trigonométriques sin nx, cos nx, où n est un nombre entier (y compris $n = 0$ pour cos nx) sur l'intervalle $(-\pi, \pi)$ par rapport au poids $\varphi(x) \equiv 1$.

Dans les applications on rencontre souvent des systèmes de *polynômes* orthogonaux. Ci-dessous nous donnons les propriétés essentielles des polynômes orthogonaux utilisés dans l'ouvrage.

Les *polynômes d'Hermite* $H_n(x)$ sont dénifis par les relations suivantes *)

$$H_n(x) = (-1)^n\, e^{\frac{x^2}{2}} \frac{d^n}{dx^n}\left(e^{-\frac{x^2}{2}}\right), \quad n = 0, 1, 2, \ldots . \tag{4}$$

En intégrant une seconde fois par parties on peut facilement montrer que

$$\int_{-\infty}^{\infty} H_m(x)\, H_n(x)\, e^{-\frac{x^2}{2}}\, dx = \begin{cases} n!\,\sqrt{2\pi}, & m = n, \\ 0, & m \neq n, \end{cases} \tag{5}$$

les polynômes d'Hermite sont donc un système de polynômes orthogonaux par rapport au poids $e^{-\frac{x^2}{2}}$ sur l'intervalle $(-\infty, \infty)$.

En vertu de la définition (4) $H_n(x)$ est un polynôme de degré n, contenant pour n pair seulement des puissances paires de x, pour n impair seulement des puissances impaires de x. Trois polynômes d'Hermite successifs quelconques sont liés par la relation de récurrence suivante

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x). \tag{6}$$

Les cinq premiers polynômes sont

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1,$$

$$H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3.$$

Il est facile d'obtenir l'expression des polynômes d'Hermite de degré plus élevé à l'aide de la formule (6).

En développant en série de Taylor la fonction $e^{-\frac{t^2}{2} + tx}$ on

*) Parfois on définit les polynômes d'Hermite comme suit:

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}\left(e^{-x^2}\right).$$

Il est évident que cette forme s'obtient à partir de (4) en remplaçant x par $x\sqrt{2}$ et en introduisant le facteur $2^{n/2}$.

obtient

$$e^{-\frac{t^2}{2}+tx}=\sum_{n=1}^{\infty}\frac{H_n(x)}{n!}t^n. \qquad (7)$$

Pour $x = 0$, on trouve à partir de (7), en développant en série la fonction $e^{-\frac{t^2}{2}}$ et en comparant les coefficients des mêmes puissances de t dans les premier et second membres:

$$H_{2n}(0) = (-1)^n (2n-1)!!, \quad H_{2n-1}(0) = 0, \qquad (8)$$

où $(2n - 1)!!$ est le produit des nombres entiers impairs jusqu'à $2n - 1$ inclus.

Les *polynômes de Laguerre* $L_n^{(\alpha)}(x)$ sont définis par les relations suivantes *)

$$L_n^{(\alpha)}(x)=\frac{(-1)^n}{n!}x^{-\alpha}e^x\frac{d^n}{dx^n}(x^{n+\alpha}e^{-x}), \qquad (9)$$

$$n=0,\ 1,\ 2,\ \ldots,\ \alpha>-1,\ x>0.$$

En intégrant par parties on obtient

$$\int_0^{\infty} L_m^{(\alpha)}(x)\,L_n^{(\alpha)}(x)\,x^{\alpha}e^{-x}\,dx=\begin{cases}\binom{n+\alpha}{n}\Gamma(\alpha+1), & m=n,\\ 0, & m\neq n,\end{cases} \qquad (10)$$

où $\Gamma(x)$ est la fonction gamma.

En vertu de (10) les polynômes de Laguerre sont un système de polynômes orthogonaux par rapport au poids $x^{\alpha}e^{-x}$ sur l'intervalle $\langle 0, \infty)$.

Trois polynômes successifs quelconques de Laguerre sont reliés par la fonction de récurrence suivante:

$$nL_n^{(\alpha)}(x) = (-x + 2n + \alpha - 1)\,L_{n-1}^{(\alpha)}(x) - $$
$$- (n + \alpha - 1)\,L_{n-2}^{(\alpha)}(x), \qquad n \geqslant 2.$$

Les quatre premiers polynômes de Laguerre sont

$$L_0^{(\alpha)}(x) = 1, \quad L_1^{(\alpha)}(x) = 1 + \alpha - x,$$
$$2L_2^{(\alpha)}(x) = (\alpha + 1)(\alpha + 2) - 2x(\alpha + 2) + x^2,$$
$$3L_3^{(\alpha)}(x) = (\alpha + 1)(\alpha + 2)(\alpha + 3) - 3x(\alpha + 2)(\alpha + 3) +$$
$$+ 3x^2(\alpha + 3) - x^3,$$

et sous forme générale

$$L_n^{(\alpha)}(x)=\sum_{k=0}^{n}\binom{n+\alpha}{k-\alpha}\frac{(-x)^k}{k!}. \qquad (11)$$

*) $L_n^0(x)$ est souvent désigné simplement $L_n(x)$.

De la même manière qu'en (7) on a le développement

$$(1-t)^{-(\alpha+1)} e^{-\frac{xt}{1-t}} = \sum_{n=0}^{\infty} L_n^{(\alpha)}(x) t^n, \quad |t|<1, \tag{12}$$

qui pour $x = 0$ donne

$$L_n^{(\alpha)}(0) = \binom{n+\alpha}{n}. \tag{13}$$

Les *polynômes de Tchébychev* (de première espèce) $T_n(x)$ sont donnés par les relations

$$T_n(x) = \cos n \arccos x = \frac{1}{2}[(x+\sqrt{x^2-1})^n + (x-\sqrt{x^2-1})^n],$$
$$n = 0, 1, 2, \ldots \tag{14}$$

Il est facile de montrer que l'on a

$$\int_{-1}^{+1} T_m(x) T_n(x) \frac{dx}{\sqrt{1-x^2}} = \begin{cases} \frac{\pi}{2}, & m=n\neq 0, \\ \pi, & m=n=0, \\ 0, & m\neq n. \end{cases} \tag{15}$$

En vertu de (15) les polynômes de Tchébychev (de première espèce) sont un système de polynômes orthogonaux par rapport au poids $\frac{1}{\sqrt{1-x^2}}$ sur l'intervalle $(-1, +1)$.

Trois polynômes successifs quelconques de Tchébychev sont liés par la relation de récurrence suivante

$$T_{n+1}(x) - 2xT_n(x) + T_{n-1}(x) = 0, \qquad n \geqslant 1. \tag{16}$$

Les cinq premiers polynômes de Tchébychev sont

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1,$$
$$T_3(x) = 4x^3 - 3x, \quad T_4(x) = 8x^4 - 8x^2 + 1.$$

En vertu de la définition (14), pour n quelconque, on a

$$T_n(1) = 1, \quad T_n(-1) = (-1)^n, \tag{17}$$
$$T_{2n}(0) = (-1)^n, \quad T_{2n+1}(0) = 0. \tag{17'}$$

ANNEXE V

Fonction hypergéométrique

L'expression générale de la fonction hypergéométrique est donnée par la série suivante

$${}_rF_s(\alpha_1, \ldots, \alpha_r, \gamma_1, \ldots, \gamma_s; x) =$$
$$= \frac{\Gamma(\gamma_1) \ldots \Gamma(\gamma_s)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_r)} \sum_{n=1}^{\infty} \frac{\Gamma(\alpha_1+n) \ldots \Gamma(\alpha_r+n)}{\Gamma(\gamma_1+n) \ldots \Gamma(\gamma_s+n)} \frac{x^n}{n!}. \tag{1}$$

Pour $r = 2$, $s = 1$ on obtient la série hypergéométrique habituelle

$$ {}_2F_1(\alpha, \beta, \gamma; x) = 1 + \frac{\alpha\beta}{\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{x^2}{2!} + $$
$$ + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2)}\frac{x^3}{3!} + \ldots \quad (2) $$

Pour $\alpha = \beta = \gamma = 1$ la série (2) devient une progression géométrique de raison x.

Dans cet ouvrage on utilise souvent un autre cas particulier de la fonction (1), la fonction hypergéométrique *dégénérée* (confluente) pour $r = s = 1$

$$ {}_1F_1(\alpha, \gamma; x) = 1 + \frac{\alpha}{\gamma}x + \frac{\alpha(\alpha+1)}{\gamma(\gamma+1)}\frac{x^2}{2!} + \frac{\alpha(\alpha+1)(\alpha+2)}{\gamma(\gamma+1)(\gamma+2)}\frac{x^3}{3!} + \ldots \quad (3) $$

Pour $x > 0$ cette fonction est liée à sa valeur pour $x < 0$ par la relation

$$ {}_1F_1(\alpha, \gamma; x) = e^x \, {}_1F_1(\gamma - \alpha, \gamma; -x). \quad (4) $$

Pour des grandes valeurs négatives de l'argument x on a le développement asymptotique suivant:

$$ {}_1F_1(\alpha, \gamma; -x) \sim \frac{\Gamma(\gamma)}{\Gamma(\gamma-\alpha)}\frac{1}{x^\alpha}\left[1 + \frac{\alpha(\alpha-\gamma+1)}{x} + \right. $$
$$ \left. + \frac{\alpha(\alpha+1)(\alpha-\gamma+1)(\alpha-\gamma+2)}{2x^2} + \ldots\right]. \quad (5) $$

Si $\alpha = -n$ (n un entier positif), ${}_1F_1(-n, \gamma; x)$ devient un polynôme de puissance n par rapport à x. Pour $\alpha = n$, $\gamma = m$ (n, m entiers), ${}_1F_1(n, m, x)$ s'exprime par des polynômes et des fonctions exponentielles de l'argument x. Si $\alpha = \frac{n}{2}$, $\gamma = m$, ${}_1F_1\left(\frac{n}{2}, m; x\right)$ se définit par des exponentielles et des fonctions de Bessel de l'argument x. Ainsi, par exemple,

$$ {}_1F_1\left(\frac{1}{2}, 1; -x\right) = e^{\frac{x}{2}} I_0\left(\frac{x}{2}\right), \quad (6) $$

$$ {}_1F_1\left(-\frac{1}{2}, 1; -x\right) = e^{-\frac{x}{2}}\left[(1+x)I_0\left(\frac{x}{2}\right) + xI_2\left(\frac{x}{2}\right)\right], \quad (6') $$

$$ {}_1F_1\left(\frac{1}{2}, 2; -x\right) = e^{-\frac{x}{2}}\left[I_0\left(\frac{x}{2}\right) + I_1\left(\frac{x}{2}\right)\right]. \quad (6'') $$

L'intégrale assez répandue du produit de fonctions polynomiale, de Bessel et exponentielle s'exprime à l'aide d'une fonction

hypergéométrique dégénérée

$$\int_0^\infty t^{\mu-1} J_\nu(\alpha t)\, e^{-\beta^2 t^2} dt = \frac{\Gamma\left(\frac{\mu+\nu}{2}\right)\left(\frac{\alpha}{2\beta}\right)^\nu}{2\beta^\mu \Gamma(\nu+1)}\, {}_1F_1\left(\frac{\mu+\nu}{2}, \nu+1; -\frac{\alpha^2}{4\beta^2}\right), \quad (7)$$

$$\alpha > 0, \ \beta > 0, \ \mu+\nu > 0.$$

En remplaçant α par $i\alpha$ [compte tenu de (4)] on obtient à partir de (7) l'expression de l'intégrale contenant une fonction de Bessel de l'argument imaginaire:

$$\int_0^\infty t^{\mu-1} I_\nu(\alpha t)\, e^{-\beta^2 t^2} dt =$$

$$= \frac{\Gamma\left(\frac{\mu+\nu}{2}\right)\left(\frac{\alpha}{2\beta}\right)^\nu}{2\beta^\mu \Gamma(\nu+1)}\, e^{\frac{\alpha^2}{4\beta^2}}\, {}_1F_1\left(\frac{\nu-\mu}{2}+1, \nu+1; -\frac{\alpha^2}{4\beta^2}\right). \quad (7')$$

On peut également exprimer à l'aide de la fonction hypergéométrique les dérivées et l'intégrale de la fonction

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \varphi^{(n)}(x) = 2^{\frac{n-1}{2}} \left\{ \frac{1}{\Gamma\left(\frac{1-n}{2}\right)}\, {}_1F_1\left(\frac{1+n}{2}, \frac{1}{2}; -\frac{x^2}{2}\right) + \right.$$

$$\left. + \frac{x\sqrt{2}}{\Gamma\left(-\frac{n}{2}\right)}\, {}_1F_1\left(\frac{2+n}{2}, \frac{3}{2}; -\frac{x^2}{2}\right) \right\}, \quad (8)$$

$${}_1F_1\left(-\frac{1}{2}, \frac{1}{2}; -\frac{x^2}{2}\right) = e^{-\frac{x^2}{2}} + x\sqrt{\frac{\pi}{2}}\,[2F(x)-1], \quad (9)$$

où $F(x)$ est la fonction de Laplace.

ANNEXE VI

Transformation de Hilbert et signal analytique

Soit $S(t)$ une fonction réelle appartenant à la classe $L^p(-\infty, \infty)$, c'est-à-dire $\int_{-\infty}^{\infty} |S(t)|^p \, dt < \infty$. Pour $p \geqslant 1$ on peut définir une fonction $\sigma(t)$ dite *conjuguée* de $S(t)$, ceci à l'aide de la transformation intégrale de Hilbert

$$\sigma(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{S(\tau)}{\tau - t} d\tau, \quad (1)$$

avec

$$S(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sigma(\tau)}{\tau - t} d\tau \quad (2)$$

[pour $t = \tau$ on prend les valeurs principales (au sens de Cauchy) des intégrales].

Si $F_s(\omega)$ est le spectre (transformée de Fourier) de la fonction $S(t)$, le spectre $F_\sigma(\omega)$ de la fonction conjuguée est

$$F_\sigma(\omega) = \int_{-\infty}^{\infty} \sigma(t)\, e^{-i\omega t}\, dt = -\frac{1}{\pi} \int_{-\infty}^{\infty} S(\tau) \int_{-\infty}^{\infty} \frac{e^{-i\omega t}}{\tau - t}\, dt\, d\tau.$$

En remplaçant t par $\tau - u$ les variables d'intégration se séparent. On a alors $F_\sigma(\omega) = F_s(\omega) \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{i\omega u}}{u} du = -F_s(\omega) \frac{2i}{\pi} \int_0^{\infty} \frac{\sin \omega u}{u} du.$

Mais comme $\int_0^{\infty} \frac{\sin \omega u}{u} du = \frac{\pi}{2} \operatorname{sign} \omega$, où sign ω indique le signe de la variable ω, on a

$$F_\sigma(\omega) = -iF_s(\omega) \operatorname{sign} \omega. \tag{3}$$

En vertu de (3) on a $|F_\sigma(\omega)| = |F_s(\omega)|$ et $\arg F_\sigma(\omega) = \arg F_s(\omega) \pm \frac{\pi}{2}$. Par exemple, la fonction conjuguée de $S(t) = A_0 \cos(\omega_0 t + \varphi)$ est $\sigma(t) = A_0 \sin(\omega_0 t + \varphi)$.

Formons sur l'axe réel t la fonction complexe $Z(t) = S(t) + i\sigma(t)$. On peut montrer que pour que la fonction complexe $Z(t)$ soit la limite, pour $u \to 0$, de la fonction analytique $Z(t + iu)$, il faut et il suffit que l'une des deux conditions suivantes soit remplie : 1) les fonctions $S(t)$ et $\sigma(t)$ sont conjuguées ; 2) la transformée de Fourier $F_z(\omega)$ de $Z(t)$ est identiquement nulle pour $\omega < 0$. Si l'une des conditions est remplie, l'autre l'est aussi.

La fonction complexe $Z(t)$ de la variable réelle t satisfaisant à l'une des conditions mentionnées est appelée *signal analytique* correspondant à $S(t)$. Désignons par $a(t)$ et $\Phi(t)$ le module et l'argument d'un signal analytique, c'est-à-dire posons

$$Z(t) = a(t)\, e^{i\Phi(t)}. \tag{4}$$

On a alors

$$S(t) = \operatorname{Re} Z(t) = a(t) \cos \Phi(t), \tag{5}$$

$$\sigma(t) = \operatorname{Im} Z(t) = a(t) \sin \Phi(t), \tag{6}$$

d'où

$$a(t) = \sqrt{S^2(t) + \sigma^2(t)}, \tag{7}$$

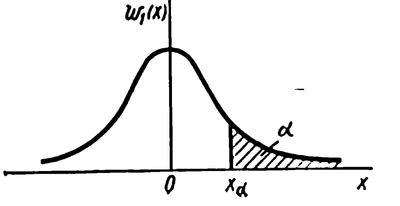
$$\Phi(t) = \operatorname{arc\,tg} \frac{\sigma(t)}{S(t)}. \tag{8}$$

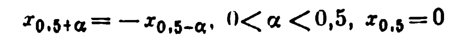
Les fonctions $a(t)$ et $\Phi(t)$ sont appelées *enveloppe* et *phase* de $S(t)$. Comme à une fonction $S(t)$ donnée, correspond d'une manière univoque le signal analytique $Z(t)$, et par conséquent, l'enveloppe $a(t) = |Z(t)|$ et la phase $\Phi(t) = \arg Z(t)$, la représentation de la fonction $S(t)$ sous la forme (5), compte tenu de (6) à (8), est *univoque*.

ANNEXE VII

Quantiles de la loi normale $w_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

α	0,0005	0,0010	0,0015	0,0020	0,0025	0,0030	0,0035	0,0040	0,0045
x_α	3,29053	3,09023	2,96774	2,87816	2,80703	2,74778	2,69684	2,65207	2,61205
α	0,005	0,006	0,007	0,008	0,009	0,010	0,015	0,020	0,025
x_α	2,57583	2,51214	2,45726	2,40892	2,36562	2,32635	2,17009	2,05375	1,95996
α	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10	0,12
x_α	1,88079	1,75069	1,64485	1,55477	1,47579	1,40507	1,34076	1,28155	1,17499
α	0,14	0,16	0,18	0,20	0,25	0,30	0,35	0,40	0,45
x_α	1,08032	0,99446	0,91537	0,84162	0,67449	0,52440	0,38532	0,25335	0,12566





$$x_{0,5+\alpha} = -x_{0,5-\alpha}, \quad 0 < \alpha < 0,5, \quad x_{0,5} = 0$$

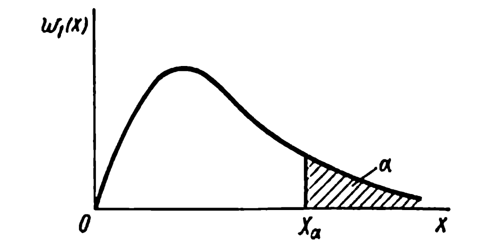
ANNEXE VIII

Quantiles de la loi du χ^2 $w_1(x) = \frac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x \geqslant 0$

n \ α	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
1	0,05157	0,04393	0,03157	0,03628	0,03982	0,00393	0,0158	0,0642	0,102	0,148	0,455
2	0,00200	0,0100	0,0201	0,0404	0,0506	0,103	0,211	0,446	0,575	0,713	1,386
3	0,0243	0,0717	0,115	0,185	0,216	0,352	0,584	1,005	1,213	1,424	2,366
4	0,0908	0,207	0,297	0,429	0,484	0,711	1,064	1,649	1,923	2,195	3,357
5	0,210	0,412	0,554	0,752	0,831	1,145	1,610	2,343	2,675	3,000	4,351
6	0,381	0,676	0,872	1,134	1,237	1,635	2,204	3,070	3,455	3,828	5,348
7	0,598	0,989	1,239	1,564	1,690	2,167	2,833	3,822	4,255	4,671	6,346
8	0,857	1,344	1,646	2,032	2,180	2,733	3,490	4,594	5,071	5,527	7,344
9	1,152	1,735	2,088	2,532	2,700	3,325	4,168	5,380	6,899	6,393	8,343
10	1,479	2,156	2,558	3,059	3,247	3,940	4,865	6,179	6,737	7,267	9,342
11	1,834	2,603	3,053	3,609	3,816	4,575	5,578	6,989	7,584	8,148	10,341
12	2,214	3,074	3,571	4,178	4,404	5,226	6,304	7,807	8,438	9,034	11,340
13	2,617	3,565	4,107	4,765	5,009	5,892	7,042	8,634	9,299	9,926	12,340
14	3,041	4,075	4,660	5,368	5,629	6,571	7,790	9,467	10,165	10,821	13,339
15	3,483	4,601	5,229	5,985	6,262	7,261	8,547	10,307	11,036	11,721	14,339
16	3,942	5,142	5,812	6,614	6,908	7,962	9,312	11,152	11,912	12,624	15,338
17	4,416	5,697	6,408	7,255	7,564	8,672	10,085	12,002	12,792	13,531	16,338
18	4,905	6,265	7,015	7,906	8,231	9,390	10,865	12,857	13,675	14,440	17,338
19	5,407	6,844	7,633	8,567	8,907	10,117	11,651	13,716	14,562	15,352	18,338
20	5,921	7,434	8,260	9,237	9,591	10,851	12,443	14,578	15,452	16,266	19,337
21	6,447	8,034	8,897	9,915	10,283	11,591	13,240	15,445	16,344	17,182	20,337
22	6,983	8,643	9,542	10,600	10,982	12,338	14,041	16,314	17,240	18,101	21,337
23	7,529	9,260	10,196	11,293	11,688	13,091	14,848	17,187	18,137	19,021	22,337

ANNEXE VIII (suite)

n \ α	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
24	8,085	9,886	10,856	11,992	12,401	13,848	15,659	18,062	19,037	19,943	23,337
25	8,649	10,520	11,524	12,697	13,120	14,611	16,473	18,940	19,939	20,867	24,337
26	9,222	11,160	12,198	13,409	13,844	15,379	17,292	19,820	20,843	21,792	25,336
27	9,803	11,808	12,879	14,125	14,573	16,151	18,114	20,703	21,749	22,719	26,336
28	10,391	12,461	13,565	14,847	15,308	16,928	18,939	21,588	22,657	23,647	27,336
29	10,986	13,121	14,256	15,574	16,047	17,708	19,768	22,475	23,567	24,577	28,336
30	11,588	13,787	14,953	16,306	16,791	18,493	20,599	23,364	24,478	25,508	29,336



n \ α	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
1	1,074	1,323	1,642	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	2,408	2,773	3,219	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	3,665	4,108	4,642	6,251	7,815	9,348	9,837	11,345	12,838	16,268
4	4,878	5,385	5,989	7,779	9,488	11,143	11,668	13,277	14,860	18,465
5	6,064	6,626	7,289	9,236	11,070	12,832	13,388	15,086	16,750	20,517

ANNEXE VIII (suite)

n \ α	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
6	7,231	7,841	8,558	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	8,383	9,037	9,803	12,017	14,067	16,013	16,622	18,475	20,278	24,322
8	9,524	10,219	11,030	13,362	15,507	17,535	18,168	20,090	21,955	26,125
9	10,656	11,389	12,242	14,684	16,919	19,023	19,679	21,666	23,589	27,877
10	11,781	12,549	13,442	15,987	18,307	20,483	21,161	23,209	25,188	29,588
11	12,899	13,701	14,631	17,275	19,675	21,920	22,618	24,725	26,757	31,264
12	14,011	14,845	15,812	18,549	21,026	23,337	24,054	26,217	28,300	32,909
13	15,119	15,984	16,985	19,812	22,362	24,736	25,472	27,688	29,819	34,528
14	16,222	17,117	18,151	21,064	23,685	26,119	26,873	29,141	31,319	36,123
15	17,322	18,245	19,311	22,307	24,996	27,488	28,259	30,578	32,801	37,679
16	18,418	19,369	20,465	23,542	26,296	28,845	29,633	32,000	34,267	39,252
17	19,511	20,489	21,615	24,769	27,587	30,191	30,995	33,409	35,718	40,790
18	20,601	21,605	22,760	25,989	28,869	31,526	32,346	34,805	37,156	42,312
19	21,689	22,718	23,900	27,204	30,144	32,852	33,687	36,191	38,582	43,820
20	22,775	23,828	25,038	28,412	31,410	34,170	35,020	37,566	39,997	45,315
21	23,858	24,935	26,171	29,615	32,671	35,479	36,343	38,932	41,401	46,797
22	24,939	26,039	27,301	30,813	33,924	36,781	37,659	40,289	42,796	48,268
23	26,018	27,141	28,429	32,007	35,172	38,076	38,968	41,638	44,181	49,728
24	27,096	28,241	29,553	33,196	36,415	39,364	40,270	42,980	45,558	51,179
25	28,172	29,339	30,675	34,382	37,652	40,646	41,566	44,314	46,928	52,620
26	29,246	30,434	31,795	35,563	38,885	41,923	42,856	45,642	48,290	54,052
27	30,319	31,528	32,912	36,741	40,113	43,194	44,140	46,963	49,645	55,476
28	31,391	32,620	34,027	37,916	41,337	44,461	45,419	48,278	50,993	56,893
29	32,461	33,711	35,139	39,087	42,557	45,722	46,693	49,588	52,336	58,302
30	33,530	34,800	36,250	40,256	43,773	46,979	47,962	50,892	53,672	59,703

Nota. Pour $n \gg 1$

$$\chi^2_\alpha \sim \frac{1}{2}(x_\alpha \pm \sqrt{2n-1})^2 ,$$

où x_α est le quantile de la loi normale (voir annexe VIII).

ANNEXE IX

Quantiles de la loi de Student

Les valeurs t_α satisfont à l'égalité $\int_{t_\alpha}^{\infty} s_n(t)\,dt = \alpha$

$n-1$ \ α	0,45	0,4	0,35	0,3	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	142	289	445	617	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,965	31,598
3	137	277	424	584	765	0,978	1,250	1,638	2,353	3,181	4,541	5,841	12,941
4	134	271	414	569	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	132	267	408	559	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	131	265	404	553	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	130	263	402	549	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	130	262	399	546	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	129	261	398	543	703	883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	129	260	397	542	700	879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
12	128	259	395	539	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
14	128	258	393	537	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
16	128	258	392	535	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
18	127	257	391	534	688	862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
20	127	257	391	533	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
25	127	256	390	531	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
30	126	256	389	530	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
60	126	254	387	527	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
∞	126	253	385	524	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

$S_n(t)$ — 0 — t_α — t — α

$0{,}5+\alpha = -t_{0{,}5-\alpha},\ 0 < \alpha < 0{,}5,\ t_{0{,}5} = 0$

ANNEXE X

Ellipsoïde de corrélation

Soit un ensemble de n variables aléatoires $\xi_1, \ldots, \xi_n$. Désignons par a_i la valeur moyenne de la variable aléatoire ξ_i et par $r_{ij} = m_1 \{(\xi_i - a_i)(\xi_j - a_j)\}$ la covariation des variables aléatoires ξ_i et ξ_j. La matrice M formée par les éléments r_{ij} $(i, j = 1, \ldots, n)$ représente la matrice corrélationnelle de l'ensemble des variables aléatoires envisagées.

L'ellipsoïde d'équation

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{D_{ij}}{D} (z_i - a_i)(z_j - a_j) = 1, \quad D > 0, \tag{1}$$

est appelé *ellipsoïde de corrélation*. Dans l'équation (1) D_{ij} sont les cofacteurs des éléments a_{ij} dans la matrice M, et D son déterminant. Par une transformation linéaire (orthogonale) des variables, la forme quadratique dans le premier membre de (1) devient une somme de carrés, soit

$$\sum_{k=1}^{n} \frac{y_k^2}{\lambda_k} = 1. \tag{2}$$

λ_k sont ici les nombres caractéristiques de la matrice de corrélation, déterminés par les racines de l'équation

$$|\mathbf{M} - \lambda \mathbf{I}| = 0,$$

où I est la matrice unité.

L'équation (2) est un ellipsoïde à n dimensions de demi-axes $\sqrt{\lambda_k}$ auquel se ramène l'ellipsoïde (1) par rotation (transformation orthogonale). Le volume de l'ellipsoïde envisagé est égal à

$$V = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} \sqrt{D}. \tag{3}$$

Ainsi, le carré du volume de l'ellipsoïde est proportionnel à la valeur du déterminant D, souvent appelée *variance généralisée*. Pour $n = 1$ cette grandeur est égale à

$$D = m_1 \{(\xi - a)^2\} = \sigma^2,$$

c'est-à-dire qu'elle coïncide avec la variance habituelle d'une variable aléatoire, et pour $n = 2$ on a

$$D = \begin{vmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\sigma_2^2(1 - r^2),$$

où σ_1^2, σ_2^2 et r sont les variances et le coefficient de corrélation des deux variables aléatoires.

Si l'on étale l'ellipsoïde de corrélation (1) de $\sqrt{n+2}$ fois on obtient l'*ellipsoïde des variances*. Pour un ensemble de variables aléatoires normalement distribuées la densité de probabilité sur l'ellipsoïde des variances est constante. Pour une loi de distribution arbitraire l'ellipsoïde de corrélation a cette particularité que l'ensemble des variables aléatoires *uniformément* réparties dans un domaine de l'espace à n dimensions, limité par cet ellipsoïde, a les mêmes premier et second moments (covariation) que l'ensemble donné des variables aléatoires.

ANNEXE XI

Régression

Soient ξ et η des grandeurs aléatoires liées caractérisées par la densité de probabilité mutuelle conjointe $w_2(x, y)$. La moyenne conditionnelle η pour $\xi = x$ [voir (2.108), t. I] considérée comme une fonction de la variable x détermine la *courbe de régression*, soit

$$z = m_1\{\eta \mid x\} = \frac{\int_{-\infty}^{\infty} y w_2(x, y)\, dy}{\int_{-\infty}^{\infty} w_2(x, y)\, dy} = \int_{-\infty}^{\infty} y w(y \mid x)\, dy. \qquad (1)$$

Il est facile de montrer (voir le problème 2.6, t. I) que parmi toutes les fonctions possibles $g(x)$, l'écart quadratique $m_1\{[\eta - g(\xi)]^2\}$ est minimal pour

$$g(x) = m_1\{\eta \mid x\}. \qquad (2)$$

Dans certains cas on envisage l'approximation de η au moyen de ξ d'après le critère de minimum du carré moyen de l'écart de η par rapport à $f(\xi)$ pour une classe donnée de fonctions $g(x)$, définies à des paramètres inconnus près. Par exemple, si $g(x) = a_1 x + a_2$, on cherche des valeurs a_1, a_2 pour lesquelles $m_1\{(\eta - a_1\xi - a_2)^2\}$ sont minimales. La droite $z = a_1 x + a_2$ est dans ce cas appelée droite de régression quadratique moyenne. Un cas plus général est celui d'une régression quadratique moyenne polynomiale quand

$$g(x) = a_1 x^n + a_2 x^{n-1} + \ldots + a_{n+1}.$$

Dans le cas d'une régression quadratique moyenne linéaire, le minimum de $m_1\{(\eta - a_1\xi - a_2)^2\}$ correspond à

$$a_1 = r\sqrt{\frac{M_2\{\eta\}}{M_2\{\xi\}}}, \quad a_2 = m_1\{\eta\} - a_1 m_1\{\xi\}, \qquad (3)$$

où r est le coefficient de corrélation des grandeurs aléatoires ξ et η. La droite de régression quadratique moyenne est alors donnée par l'équation

$$z = m_1\{\eta\} + [x - m_1\{\xi\}]\,\frac{M_2\{\xi\eta\}}{M_2\{\xi\}}. \tag{4}$$

Si ξ et η sont des grandeurs aléatoires normales liées, le second membre de (4) coïncide exactement avec la moyenne conditionnelle de η pour $\xi = x$ (voir t. I, page 78). Par conséquent, la droite de régression quadratique moyenne pour des grandeurs aléatoires normales coïncide avec la courbe de régression. Autrement dit, l'approximation linéaire de η par ξ suivant le critère de minimum du carré moyen de l'erreur est la meilleure.

Les notions introduites ci-dessus peuvent être généralisées à un ensemble fini quelconque de grandeurs aléatoires interdépendantes $\xi_1, \ldots, \xi_n$. La valeur moyenne conditionnelle de ξ_1 pour $\xi_i = x_i$, $i = 2, \ldots, n$ est égale à

$$m_1\{\xi_1 \mid x_2, \ldots, x_n\} = \frac{\int\limits_{-\infty}^{\infty} x_1 w_n(x_1, \ldots, x_n)\, dx_1}{\int\limits_{-\infty}^{\infty} w_n(x_1, \ldots, x_n)\, dx_1}. \tag{5}$$

Le lieu géométrique des points $(m_1\{.\}, x_2, \ldots, x_n)$ pour toutes les valeurs $x_2, \ldots, x_n$ est l'*hypersurface de régression*. On peut définir l'*hyperplan de la régression quadratique moyenne*

$$z_1 = m_1\{\xi_1\} + \sum_{i=2}^{n} \rho_{1i}\,[x_i - m_1\{\xi_i\}]. \tag{6}$$

Les grandeurs ρ_{1i} sont données par le système d'équations linéaires

$$\sum_{j=2}^{n} b_{ij}\rho_{1j} = b_{i1}, \quad i = 2, \ldots, n, \tag{7}$$

où

$$b_{ij} = m_1\{\xi_i\xi_j\}, \tag{7'}$$

de sorte qu'en désignant par D_{ij} les cofacteurs dans la matrice $\| b_{ij} \|$ on a

$$\rho_{1i} = \frac{D_{1i}}{D_{11}}. \tag{8}$$

On peut montrer que l'hypersurface de régression pour un ensemble de grandeurs aléatoires *normalement* distribuées coïncide avec l'hyperplan de la régression quadratique moyenne.

BIBLIOGRAPHIE

CHAPITRE 1

1. A. E. BACHARINOV, B. S. FLEISCHMAN — Méthodes d'échantillonnage progressif et applications (en russe). Ed. « Sovetskoe radio», 1962, chap. 1, 3.
2. I. A. BOLCHAKOV, L. S. GOUTKINE, B. R. LÉVINE, R. L. STRATONOVITCH — Bases mathématiques de la radiotechnique moderne (en russe). Ed. « Sovetskoe radio », 1968, chap. 6.
3. A. WALD — Echantillonnage progressif (en russe). Ed. « Physmathguiz », 1960, chap. 1 à 4.
4. W. B. DAVENPORT, W. L. ROOT — An introduction to the theory of random signals and noise. Mc Graw-Hill, N. Y., 1958, § 14.2.
5. I. V. DOUNINE-BARKOVSKI, N. V. SMIRNOV — Cours de théorie des probabilités et de statistique mathématique (en russe). Ed. « Nauka », 1965, chap. VII.
6. H. CRAMER — Mathematical methods of statistics. Princeton Univ. press, Princeton, 1946.
7. E. L. LEHMANN — Testing statistical hypothesis. John Wiley, New York, 1959.
8. J. V. LINNIK — Problèmes statistiques à paramètres brouilleurs (en russe). Ed. « Nauka », 1966, chap. III, IV.
9. D. MIDDLETON — An introduction to statistical communication theory. Mc Graw-Hill, N.Y., 1960, chap. 18.
10. D. MIDDLETON — Topics on communication theory. Mc Graw-Hill, N.Y., 1964, chap. 1, 2.
11. E. M. HASEN, Z. M. BENENSON — Tests d'échantillonnage progressif appliqués à quelques problèmes de distinction de plusieurs hypothèses ou d'hypothèses composites (en russe). Thèses des rapports au Congrès international des mathématiciens, section 11, Moscou, 1966.
12. C. W. HELSTROM — Statistical theory of signal detection. Pergamon press, 1960, chap. 3.
13. C. E. SHANNON — A mathematical theory of communication. Bell Syst. Techn. Journ., vol. 27, 1948.
14. A. WALD — Statistical decision functions. John Wiley, New York, 1950, chap. 1, 3.

CHAPITRE 2

Monographies

1. T. W. ANDERSON — An introduction to multivariate statistical analysis. John Wiley, New York, 1958, chap. 2, 3.
2. B. L. VAN DER WAERDEN — Mathematische statistik. Berlin, 1957, chap. 8.
3. I. V. DOUNINE-BARKOVSKI, N. V. SMIRNOV — Cours de théorie des probabilités et de statistique mathématique (en russe). Ed. « Nauka », 1965, chap. 6, 7.

1/2 27—0682

4. B.V. GNEDENKO — The theory of probability. Chelsea Publishing Co, N.Y., 1962, chap. 11.
5. H. CRAMER — Mathematical methods of statistics. Princeton Univ. press, Princeton, 1946, chap. 3.
6. E. L. LEHMANN — Testing statistical hypothesis. John Wiley, New York, 1959, chap. 1.
7. J. V. LINNIK — Méthode des moindres carrés et fondements de la théorie du traitement des observations (en russe). Physmathguiz, 1958, chap. 3, 4.
8. J. V. LINNIK — Problèmes statistiques à paramètres brouilleurs (en russe). Ed. « Nauka », 1966, chap. 2, 3, 7.
9. D. MIDDLETON — An introduction to statistical communication theory. Mc Graw-Hill, N.Y., 1960, chap. 21.
10. C. W. HELSTROM — Statistical theory of signal detection. Pergamon press, 1960, chap. 7.

Articles

11. L. N. BOLCHEV — Précision de l'inégalité de Cramer-Rao (en russe). Recueil « Théoria véroïatnostei i eïe priménénia », 1961, sér. 3.
12. L. N. BOLCHEV — Transformations asymptotiques de Pearson (en russe). Recueil « Théoria véroïatnostei i eïe priménénia », 1963, sér. 2.
13. E. B. DYNKINE — Statistiques nécessaires et suffisantes pour une famille de répartitions des probabilités (en russe). Recueil «Uspekhi mathematitcheskikh nauk », 1951, sér. 1.
14. I. I. GHIKHMAN — Sur un critère non paramétrique d'homogénéité de k échantillons (en russe). Recueil « Théoria véroïatnostei i eïe priménénia », 1957, sér. 3.
15. E. J. NADARAÏA — Nouvelles estimations des répartitions (en russe). Recueil « Théoria véroïatnostei i eïe priménénia », 1964, sér. 3.
16. H. E. DANIELS — The asymptotic efficiency of a maximum likelyhood estimator. Proc. 4. Berkl. Symp. on Math. Stat. and Probab., Univ. Calif. press, 1961.
17. E. PARZEN — On estimation of a probability density function and mode. Ann. Math. Statist., 1962, v. 33, nº 3.

CHAPITRE 3

Monographies

1. L. A. VAINSTEIN, V. D. ZOUBAKOV — Sélection des signaux noyés dans les bruits (en russe). Ed. « Sovetskoe radio », 1959, annexe III.
2. U. GRENANDER — Stochastic processes and statistical inferences. Arkiv. mat. 1 (1950), nº 17, p. 195 à 277.
3. W. B. DAVENPORT, W. L. ROOT — An introduction to the theory of random signals and noise. Mc Graw-Hill, N.Y., 1958, chap. 14 et annexe 2.
4. J. L. DOOB — Stochastic processes. John Wiley, N.Y., 1953, chap. 7, § 9.
5. D. MIDDLETON — An introduction to statistical communication theory. Mc Graw-Hill, N.Y., 1960, vol. I, chap. 4; vol. II, chap. 16.
6. M. G. SÉRÉBRIANNIKOV, A. L. PERVOZVANSKI — Discernement des périodicités cachées (en russe). Ed. « Nauka », 1965, chap. 5.
7. E. J. HANNAN — Time series analysis. John Wiley, N.Y., 1962.
8. A. A. KHARKÉVITCH — Spectres et analyse (en russe). Ed. Gostechizdat, 1949.

Articles

9. J. K. BELAEV — Processus aléatoires analytiques (en russe). Recueil « Théoria véroïatnostei i eïe priménénia », 1959, sér. 4.
10. J. GAEK — Estimation linéaire de la moyenne d'un processus stationnaire à fonction de corrélation convexe. Journal de mathématiques tchèque, 1956, sér. 6.

11. B. R. LÉVINE, J. S. CHINAKOV — Forme limite de l'estimation de Bayes des coefficients de régression pour un bruit normal non stationnaire (en russe). Recueil « Problemy pérédatchi informatsii », 1967, nº 1.
12. B. F. PISSARENKO — Sur un problème de détection d'un processus aléatoire noyé dans le bruit (en russe). Journ. « Radiotechnika i electronica », 1961, nº 4.
13. B. F. PISSARENKO — Sur le calcul du rapport de vraisemblance des processus gaussiens à spectre rationnel (en russe). Recueil « Théoria véroïatnostei i eïe prilogénia », 1965, sér. 2.
14. R. B. BLACKMAN, I. W. TUKEY — The measurement of power spectre from point of view of communication engineering. BSTJ, 1958, nºs, 1, 2.
15. U. GRENANDER, M. ROSENBLATT — Some problems in estimating the spectrum of time series. Proc. Third Berkley Symp. Math. Stat., Los Angeles, 1956, v. 2.
16. E. J. KELLY, I. S. REED, W. L. ROOT — The detection of radar echoes in noise. I. J. Soc. Ind. Appl. Math. 1960, v. 8, nº 2.
17. E. M. HOFSTETTER — Some results on the stochastic signal parameter estimation problem. IEEE Trans., 1965, IT-11, nº 3.
18. M. J. LÉVIN — Pover spectrum parameter estimation. IEEE Trans., 1965, IT-11, nº 1.
19. E. REICH, P. SWERLING — Detection of sine wave in gaussian noise. J. Appl. Phys. Rev., 1953, v. 24, nº 3.
20. D. SLEPIAN — Some comments on the detection of gaussian signals in gaussian noise. IRE Trans., 1958, IT-4, nº 2.
21. T. KAILATH — Some results on singular detection. Information and control, 1966, v. 9, april.

CHAPITRE 4

Monographies

1. J. S. BENDAT — Principles and applications of random noise theory. John Wiley, N.Y., 1958, chap. 4.
2. L. A. VAINSTEIN, V. D. ZOUBAKOV — Sélection des signaux noyés dans le bruit (en russe). Ed. « Sovetskoe radio », 1959, chap. 2, 3.
3. N. WIENER — Nonlinear problems in random theory. John Wiley, N. Y., 1958.
4. W. B. DAVENPORT, W. L. ROOT — An introduction to the theory of random signals and noise. Mc Graw-Hill, N.Y., 1958, chap. 11.
5. R. DEUTSCH — Nonlinear transformations of random processes. Prentice-Hall, 1962, chap. 6.
6. J. S. LÉSINE — Filtres optimaux et accumulateurs de séquences impulsionnelles (en russe). Ed. « Sovetskoe radio ». 1963, chap. 1, 2.
7. R. LEE — Optimal estimation, identification and control. The MIT press, Cambridge, 1964, chap. 3.
8. J. H. LANNING, R. H. BATTIN — Random processes in automatic control. Mc Graw-Hill, N.Y., 1956, chap. 7, 8.
9. D. MIDDLETON — An introduction to statistical communication theory. Mc Graw-Hill, N.Y., 1960, chap. 16.
10. D. MIDDLETON — Topics on communication theory. Mc Graw-Hill, N. Y., 1964, chap. 3.
11. V. S. POUGATCHEV — Théorie des fonctions aléatoires et applications aux problèmes de régulation automatique (en russe). Ed. « Physmathguiz », 1962, chap. 16 à 18.
12. V. V. SOLODOVNIKOV — Dynamique statistique des systèmes linéaires de régulation automatique (en russe). Ed. « Physmathguiz », 1960, chap. 7, 8.
13. R. L. STRATONOVITCH — Processus markoviens conditionnels et applica-

tions à la théorie de la régulation automatique (en russe). Ed. de l'Université de Moscou, 1966.
14. V. I. TIKHONOV — Radiotechnique statistique (en russe). Ed. « Sovetskoe radio », 1966, chap. 13.
15. A. A. KHARKÉVITCH — Protection contre les bruits (en russe). Ed. « Physmathguiz », 1963, §§ 10, 11.
16. J. F. BARRETT — Application of the theory of functionals to communication problems. Cambr. Univ., 1955.
17. N. WIENER — The interpolation, extrapolation and smoothing of stationary time series. Jonh Wiley, N.Y., 1949.

Articles

18. H. W. BODE, C. E. SHANNON — A simplified derivation of linear least square smoothing and prediction theory. PIRE 38, 4, 417-425, 1950.
19. J. KATZNELSON, L. GOOD — Construction of nonlinear filters and control systems. Dans le livre de H. L. VAN TREES — Synthesis of optimum nonlinear control systems. Cambridge. The MIT press, 1962.
20. A. N. KOLMOGOROV — Interpolation et extrapolation des séquences aléatoires stationnaires (en russe). Journ. « Izvestia Akad. Nauk SSSR », sér. math., 1941, nº 5.
21. R. L. STRATONOVITCH — Application des processus de Markov à la filtration optimale des signaux (en russe). Journ. « Radiotechnica i électronica », 1960, nº 11.
22. I. G. KHANOVITCH, G. G. YANOVSKI — Méthodes permettant de préciser les prédictions linéaires des séries temporelles stationnaires (en russe). XXIIe Session scientifique consacrée au Jour de la radio. Section de la théorie de l'information, 1966.
23. L. A. ZADEH — On the representation of nonlinear operation. IRE Wescon Conv. Rec., 1957, pt. 2.
24. L. A. ZADEH — Progress in information theory in USA. IRE Trans., 1961, IT-7, nº 2, 1963, IT-9, nº 4.

CHAPITRE 5

Monographies

1. L. A. VAINSTEIN, V. D. ZOUBAKOV — Sélection des signaux noyés dans le bruit (en russe). Ed. « Sovetskoe radio », 1960, chap. 1 à 3.
2. W. B. DAVENPORT, W. L. ROOT — An introduction to the theory of random signals and noise. Mc Graw-Hill, N.Y., 1958, chap. 14.
3. L. S. GOUTKINE — Théorie des méthodes optimales de réception radio en présence de bruits fluctuationnels (en russe). Ed. « Gossénergoïzdat », 1961, chap. 9, 10.
4. V. A. KOTELNIKOV — Théorie de la stabilité potentielle aux bruits (en russe). Ed. « Cossénergoïzdat », 1956.
5. B. R. LÉVINE — Théorie des processus aléatoires et applications radiotechniques (en russe), 2me éd. Ed. « Sovetskoe radio », 1960, chap. 11.
6. D. MIDDLETON — An introduction to statistical communication theory. Mc Graw-Hill, N.Y., 1960, chap. 19, 20.
7. D. MIDDLETON — Topics on communication theory. Mc Graw-Hill, N.Y., 1964, chap. 2.
8. L. M. FINK — Théorie de la transmission des communications discrètes (en russe). Ed. «Sovetskoe radio », 1963, chap. 4.
9. A. A. KHARKÉVITCH — Systèmes antiparasites (en russe). Ed. « Physmathguiz», 1964, §§ 8 à 14.
10. V. I. TIKHONOV — Radiotechnique statistique (en russe). Ed. « Sovetskoe radio », 1966, chap. 11.

11. C. W. HELSTROM — Statistical theory of signal detection. Pergamon Press. 1960, chap. 5.
12. J. D. SHIRMAN, V. N. GOLIKOV — Fondements de la théorie de la détection des signaux radar et mesure de leurs paramètres (en russe). Ed. « Sovetskoe radio », chap. 4 à 6.
13. M. SCHWARTZ, W. R. BENNETT, S. STEIN — Communication system and techniques. Mc Graw-Hill, N. Y., 1966, pt. 1.
14. I. SELIN — Detection theory. New Jersey, Princeton Univ. press, 1965.
15. A. WALD — Statistical decision functions. N. Y., 1950.
16. J. V. VITERBI — Principles of coherent communication. Mc Graw-Hill, N.Y., 1966.

Articles

17. B. R. LÉVINE — Méthodes optimales de phase de détection des signaux (en russe). Journ. « Radiotechnica i électronica », 1960, nº 4.
18. W. W. PETERSON, T. G. BIRDSALL, W. C. FOX — The theory of signal detectability. Trans. IRE PGIT-4, 171-212, 1954.
19. R. ESPOSITO, D. MIDDLETON, J. A. MULLEN — Advantages of signals subject to slow Rayleigh fading. IEEE Trans., 1965, IT-11, nº 4.
20. E. J. KELLY, I. S. REED, W. L. ROOT — The detection of radar echoes in noise. I. J. Soc. Ind. Appl. Math., 1960, v. 8, nº 2.
21. T. T. KADOTA — Optimum reception of binary gaussian signals. BSTJ, 1964, Nov., v. 43.
22. T. T. KADOTA — Simultaneous orthogonal expansion of two stationary gaussian processes. BSTJ, 1966, sept., v. 45.
23. J. B. TCHERNIAK — Propriétés linéaires des circuits limiteur-filtre à large bande (en russe). Journ. « Radiotechnica i électronica », 1962, nº 7.

CHAPITRE 6

Monographies

1. D. E. VAKMAN — Signaux compliqués et principe d'incertitude en radar (en russe). Ed. « Sovetskoe radio », 1965, chap. 2, 3.
2. P. M. WOODWARD — Probability and information theory, with application to radar. Pergamon press, 1954, chap. 5, 6, 7.
3. G. P. TARTAKOVSKI (red.) — Questions de la théorie statistique du radar, vol. 2 (en russe). Ed. « Sovetskoe radio », 1964.
4. L. S. GOUTKINE — Théorie des méthodes optimales de réception radio en présence de bruits de fluctuation (en russe). Ed. « Gossénergoizdat », 1961, chap. 13, 16.
5. D. MIDDLETON — An introduction to statistical communication theory. Mc Graw-Hill, N.Y., 1960, chap. 21.
6. D. MIDDLETON — Topics on communication theory. Mc Graw-Hill, N.Y., chap. 3.
7. S. E. FALKOVITCH — Réception des signaux radar noyés dans les bruits de fluctuation (en russe). Ed. « Sovetskoe radio », 1961, chap. 4, 5.
8. A. A. KHARKÉVITCH — Systèmes antiparasites (en russe). Ed. « Nauka », 1965, §§ 15, 16.
9. C. W. HELSTROM — Statistical theory of signal detection. Pergamon press, 1960, chap. 8.
10. J. D. CHIRMAN, V. N. GOLIKOV — Fondements de la théorie de la détection radar et mesure de leurs paramètres (en russe). Ed. « Sovetskoe radio », 1963, chap. 6, 7.
11. I. A. BOLCHAKOV — Sélection des signaux noyés dans les bruits (en russe). Ed. « Sovetskoe radio », 1969.

Articles

12. E. J. KELLY — Mesures de la distance, de la vitesse et de l'accélération en radar (version russe). Journ. « Zaroubejnaïa radioélectronica », 1962, nº 2, pp. 35 à 47.
13. B. R. LÉVINE, J. S. CHINAKOV — Forme limite de l'estimation de Bayes des coefficients de régression pour un bruit normal non stationnaire (en russe). Journ. « Problemy pérédatchi informatsii », 1967, nº 1.
14. B. R. LÉVINE, J. S. CHINAKOV — Propriétés asymptotiques des estimations de Bayes du coefficient de régression en présence d'un bruit non stationnaire (en russe). Journ. « Problemy pérédatchi informatsii », 1967, nº 3.
15. J. V. ABBATE, D. L. SCHILLING — Estimation of random phase — and frequency modulating signals using a Bayes estimator. IEEE Trans., 1965, IT-11, nº 3.
16. E. J. KELLY, I. S. REED, W. L. ROOT — The detection of radar echoes in noise. II. J. Soc. Ind. Appl. Math., 1960, v. 8, nº 3.
17. W. M. SIBERT — A radar detection philosophy. IRE Trans., 1965, Sept. IT-2.

CHAPITRE 7

Monographies

1. T. W. ANDERSON — An introduction to multivariate statistical analysis. John Wiley, New York, 1958.
2. D. BRAVERMAN — Theory and applications. Dans « Advances in communication systems ». Academic press, New York and London, 1965.
3. R. KOULIKOVSKI — Processus optimaux et adaptatifs dans les systèmes de régulation automatique (version russe). Ed. «Nauka », 1967.
4. G. S. SEBESTYEN — Decision-making processes in pattern recognition. New York-London, Mcmillan, 1962.

Articles

5. M. A. AÏZERMAN, E. M. BRAVERMAN, L. I. ROZONOER — Fondements théoriques de la méthode des fonctions potentielles pour le problème de l'apprentissage des automates à la classification des situations d'entrée (en russe). Journ. « Avtomatica i téléméchanica », 1964, nº 6.
6. M. A. AÏZERMAN, E. M. BRAVERMAN, V. M. GLOUCHKOV, V. A. KOVALEVSKI, A. A. LETITCHEVSKI — Théorie de la reconnaissance des images et des systèmes d'apprentissage (en russe). Journ. « Izvestia Akad. nauk SSSR », Cybernétique technique, 1963, nº 5.
7. M. A. AÏZERMAN, E. M. BRAVERMAN, L. I. ROZONOER — Méthode des fonctions potentielles dans le problème de la restitution des caractéristiques des transformateurs fonctionnels d'après les points aléatoires observés (en russe). Journ. « Avtomatica i téléméchanica », 1965, nº 5.
8. M. A. AÏZERMAN, E. M. BRAVERMAN, L. I. ROZONOER — Processus de Robins-Monroe et méthodes des fonctions potentielles (en russe). Journ. « Avtomatica i téléméchanica », 1965, nº 11.
9. A. E. BACHARINOV — Procédures asymptotiques extrémales de reconnaissance des images (en russe. Journ. « Radiotechnica i électronica », 1965, n° 5.
10. S. N. BERNSTEIN — Probabilités de confiance de Fisher (en russe). Journ. « Izvestia Acad. nauk SSSR », sér. mathématique, 1941, nº 5.
11. E. B. DYNKINE — Statistiques nécessaires et suffisantes pour une famille de répartitions des probabilités (en russe). Journ. « Uspekhi mathematitcheskikh nauk », 1951.
12. V. A. LIKHAREV, V. P. TÉBIAKINE — Méthodes d'apprentissage des

systèmes de solution des problèmes de reconnaissance des images et détection des signaux noyés dans les bruits (revue des art.). Journ. « Zaroubejnaïa radioelectronica », 1967, n° 11.
13. N. V. LOGUINOV — Méthodes d'approximation stochastique (en russe). Journ. « Avtomatica i téléméchanica », 1966, n° 4.
14. J. NEUMANN — Two break-throughs in the theory of statistical decision making. Revue de l'Inst. Intern. de Stat., 30, n° 1 (1962), PU-27.
15. V. S. POUGATCHEV — Algorithmes optimaux d'apprentissage des systèmes automatiques par un maître imparfait (en russe). Journ. « Doklady Akadémii nauk SSSR», 1967, t. 172, n° 5.
16. H. ROBBINS — An emperical Bayes approach to statistics. Proceedings of the third Berkeley Simposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, Univ. Calif. press, v. I (1956), 157-163.
17. H. ROBBINS — The emperical Bayes approach to statistical decision problems. AMS, 35, n° 1 (1964).
18. A. A. KHARKÉVITCH — Choix des signes lors de la reconnaissance par les calculatrices (en russe). Journ. « Technitcheskaïa kibernetica », 1963, n° 2.
19. J. Z. TSYPKINE — Restitution des caractéristiques d'un transformateur fonctionnel d'après des points aléatoires observés (en russe). Journ. « Avtomatica i téléméchanica », 1965 n° 11.
20. J. Z. TSYPKINE — Adaptation, apprentissage et auto-apprentissage dans les systèmes automatiques (en russe). Journ. « Avtomatica i téléméchanica », 1966, n° 1.
21. S. K. TCHOW — An optimum character recognition system using decision functions. IRE Trans., EC-6, n° 4 (1957), 247-254.
22. N. ABRAMSON, D. BRAVERMAN — Learning to recognize patterns in a random environment. IRE Trans., 1962, IT-8, n° 5.
23. N. ABRAMSON, D. BRAVERMAN, G. SEBESTYEN — Pattern recognition and machine learning. IEEE Trans., 1963, IT-9, n° 4.
24. T. W. ANDERSON — Classification by multivariate analysis. Psychometrika, 1951, v. 16.
25. T. W. ANDERSON, R. R. BACHADUR — Classification into two multivariate normal distribution with different covariance matrices. Annals of mathem. stat., 1962, n° 2.
26. D. B. COOPER — Adaptive pattern recognition and signal detection using stochastic approximation. IEEE Trans., 1964, EC-13, n° 13.
27. D. B. COOPER, P. W. COOPER — Adaptive pattern recognition and signal detection without supervision. IEEE Int. Conv. Rec., 1964, p. 1.
28. S. C. FRALICK — Learning to recognize pattern without a teacher. IEEE Trans.. 1967, IT-13, n° 1.
29. E. M. GLASER — Signal detection by adaptive filters. IRE Trans., 1961, IT-7, n° 2.
30. W. H. HIGHLEYMAN — Linear decision function with application to pattern recognition. Proc. IRE, 1962, n° 6.
31. M. J. HINICH — A model for a self-adapting filter. Inform. and Control, 1962, Sept.
32. S. JOHN — On some classification statistics. Indian journ. of statistics, 1960, v. 22, p. 3-4.
33. S. JOHN — Errors in discrimination. Annals of mathem. statistics, 1961, n° 4.
34. D. G. KABE — Some results on the distribution of two random matrices used in classification procedures. Annals of math. statistics, 1963, n° 1.
35. M. KAC — A note on learning signal detection. IRE Trans., 1962, IT-8, n° 2.
36. R. L. KASHYAP, C. C. BLAYDON — Recovery of function from noisy measurements taken at randomly selected points and its application to pattern classification. Proc. IEEE, 1966, n° 8.

37. D. G. KEEHN — A note on learning for gaussian properties. IEEE Trans., 1965, IT-11, n° 1.
38. T. MARILL, D. GREEN — Effectiveness of receptors in recognition system. IEEE Trans., 1963, IT-9, n° 1.
39. M. OKATOMO — Discrimination for variance matrices. Annals of mathem. statistics, 1961, n° 1-2.
40. E. A. PATRICK, J. C. HANCOCK — The non-supervised learning of probability spaces and recognition of patterns. IEEE Int. Conv. Rec., 1965, p. VII.
41. H. J. SCUDDER — Adaptive communication receviers. IEEE Trans., 1965, IT-11, n° 2.
42. G. SEBESTYEN — Pattern recognition by an adaptive process of sample set construction. IRE Trans., 1962, IT-8, n° 3.
43. J. SPRAGINS —A note of the iterative application of Bayes rule. IEEE Trans., 1965, IT-11, n° 4.
44. J. SPRAGINS — Learning without a teacher. IEEE Trans., 1966, IT-22, n° 2.
45. A. WALD — On a statistical problem arising in the classification of an individual in one of two groups. Annals of mathem. statist., 1944, v. 15.

NOTATIONS UTILISÉES

A, A_0 — amplitude d'un signal
$A(\omega)$ — spectre d'amplitude
$A(t)$ — composante en quadrature d'un processus
a — valeur moyenne d'une variable aléatoire, amplitude d'un signal déterministe
$B(t, y)$ — fonction de corrélation d'un processus aléatoire
$B(\tau)$ — fonction de corrélation d'un processus stationnaire au sens général
$B_{\xi\eta}(t, y)$ — fonction de corrélation mutuelle des processus aléatoires $\xi(t)$ et $\eta(t)$
$B_{\xi\eta}(\tau)$ — fonction de corrélation mutuelle de deux processus $\xi(t)$ et $\eta(t)$ stationnairement liés au sens général
b_n — biais d'une estimation
$C(t)$ — composante en quadrature d'un processus
c, c^* — seuil
$C(\omega)$ — réponse en fréquence (caractéristique) d'un système linéaire
D — déterminant de la matrice de corrélation
d_N, d_T — paramètres des caractéristiques de détection d'un signal
$E(t)$ — enveloppe d'un processus aléatoire
E — énergie d'un signal
e — efficacité relative d'une estimation
$F_N(x_1, \ldots, x_N)$ — fonction de répartition à N dimensions
$F(x)$ — intégrale de Laplace
$F(\omega)$ — spectre énergétique d'un processus aléatoire, stationnaire au sens général
$F_{\xi\eta}(\omega)$ — spectre énergétique mutuel des processus aléatoires $\xi(t)$ et $\eta(t)$ stationnairement liés au sens général
$f\{\Pi\}$ — critère de qualité
G — domaine de l'espace des échantillons
H — hypothèse
$H(s)$ — entropie
$H(s \mid \gamma)$ — entropie conditionnelle
$h(t, y, u)$ — résolvante
$h(u, v)$ — fonction de transition impulsionnelle d'un système linéaire, fonction de pondération
$h^*(u, v)$ — fonction de transition impulsionnelle d'un système linéaire optimal
I_n — information de Fisher
I — quantité d'information
J_ν — fonction de Bessel de 1re espèce d'ordre ν
I_ν — fonction de Bessel de 1re espèce d'ordre ν de l'argument imaginaire

$K(u, v)$ — noyau d'une forme intégrale quadratique, caractéristiques d'un filtre du second degré
$k^*(i\omega)$ — fonction de transfert d'un système linéaire optimal
$k(i\omega)$ — fonction de transfert d'un système linéaire
$\mathrm{k}(z)$ — fonction de répartition de Kolmogorov
K — seuil
L — fonction de vraisemblance
l — rapport de vraisemblance
m^2 — moment deux de l'amplitude aléatoire d'un signal
M_k — moment d'ordre k centré
M_2 — variance
M_2^* — variance échantillonnée
m_k — moment d'ordre k
m_1 — valeur moyenne
m_1^* — moyenne échantillonnée
m — nombre d'états, nombre de domaines disjoints de l'espace des échantillons
N — dimension d'une loi multidimensionnelle
N_0 — densité spectrale (intensité) d'un bruit blanc
n — taille d'un échantillon
P — probabilité d'un événement
p — probabilité a priori
Π — fonction de pertes
$Q_n(x)$ — polynôme formé par un ensemble de polynômes orthogonaux
q — probabilité a priori
$r(t)$ — réalisation de l'enveloppe
$r(\vartheta)$ — fonction condionnelle de risque
R — risque moyen
R^* — risque de Bayes
$R(\tau)$ — coefficient de corrélation
S — espace des paramètres
$s(t)$ — processus déterministe (signal)
s — paramètre d'une distribution
s_k — états possibles
$s_n(x)$ — densité de probabilité de Student
T — durée d'une observation
t — coordonnée temporelle, erreur normée d'une estimation
U — nombre d'inversions
$u(t)$ — composante en quadrature d'un signal déterministe
$V(t)$ — solution d'une équation intégrale non homogène linéaire
$v(t)$ — composante en quadrature d'un signal déterministe
$\mathrm{W}_1(x)$, $w_1(x)$, $W_1(x)$ — densités de probabilité unidimensionnelles d'une variable aléatoire
$w_N(x_1, \ldots, x_N)$ — densité de probabilité multidimensionnelle d'un ensemble de variables aléatoires
$W_N(x_1, \ldots, x_N)$ — fonction de vraisemblance d'un échantillon de taille n
$\mathbf{X}$ — vecteur des valeurs échantillonnées
x_α — quantile
x_i — valeur échantillonnée
$x(t)$ — réalisation d'un processus aléatoire
$z(t)$ — enveloppe complexe
$Z_T(i\omega)$ — spectre d'une réalisation tronquée d'un processus aléatoire
z — variable complexe

$\bar{z}$ — grandeur imaginaire conjuguée de z
α — probabilité conditionnelle d'erreur de première espèce, probabilité conditionnelle de fausse alarme
β — probabilité conditionnelle d'erreur de seconde espèce, probabilité conditionnelle de perte d'un signal
$\Gamma(x)$ — fonction gamma complète
$\Gamma(x, y)$ — fonction gamma incomplète
γ — décision; coefficient de confiance
$\delta(x)$ — fonction delta
$\delta(\gamma \mid \mathbf{x})$ — stratégie
Δ — largeur de bande
ε — longueur relative de l'intervalle de confiance
ε^2 — erreur quadratique moyenne
Λ — fonction de vraisemblance centrée
λ_k — nombre caractéristique d'une équation intégrale linéaire
μ — rapport des probabilités a priori; paramètre d'un processus de Markov
Ψ — fonction d'incertitude
$\psi(t)$ — réalisation de la phase
ξ — variable aléatoire
$\xi(t)$ — processus aléatoire
η — variable aléatoire
$\eta(t)$ — processus aléatoire
χ^2 — variable aléatoire distribuée suivant la loi du χ^2
Im — partie imaginaire
Re — partie réelle
$\binom{n}{k}$ — nombre de combinaisons de n éléments k à k
$\in$ — symbole désignant l'appartenance à un ensemble

INDEX ALPHABÉTIQUE DES MATIÈRES

INDEX DES NOMS

TABLE DES MATIÈRES

CHAPITRE 2

STATISTIQUE DES VARIABLES ALÉATOIRES

CHAPITRE 3

STATISTIQUE DES PROCESSUS ALÉATOIRES

CHAPITRE 4

FILTRAGE DES PROCESSUS ALÉATOIRES

CHAPITRE 5

DÉTECTION DES SIGNAUX NOYÉS DANS LES BRUITS

CHAPITRE 6

SÉPARATION DES SIGNAUX NOYÉS DANS UN BRUIT

CHAPITRE 7

ÉLÉMENTS DE THÉORIE DE LA CLASSIFICATION AVEC APPRENTISSAGE

A NOS LECTEURS

Les Editions Mir vous seraient très reconnaissantes de bien vouloir leur communiquer votre opinion sur le contenu de ce livre, sa traduction et sa présentation, ainsi que toute autre suggestion.

Notre adresse: Editions Mir,
2, Pervi Rijski péréoulok,
Moscou, I-110, GSP, U.R.S.S.

Imprimé en Union Soviétique